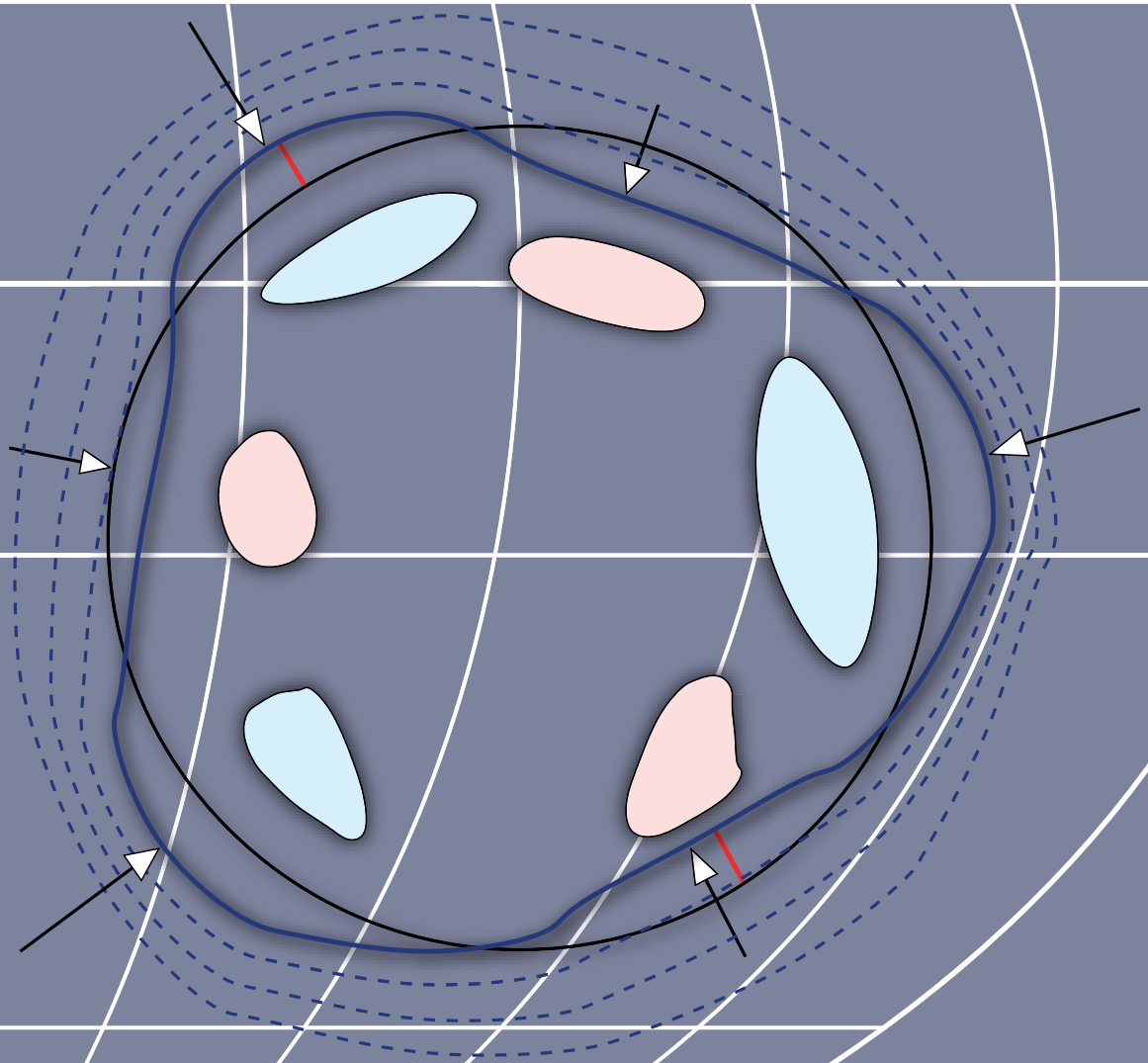
Martin Vermeer

PHYSICAL GEODESY



Aalto University

Aalto University publication series

SCIENCE + TECHNOLOGY 2/2020

Physical geodesy

Martin Vermeer

Aalto University
School of Engineering
Department of Built Environment

Aalto University publication series
SCIENCE + TECHNOLOGY 2/2020

ISBN (pdf) 978-952-60-8940-9
ISSN 1799-490X (pdf)
http://urn.fi/URN:ISBN:978-952-60-8940-9

Graphic design: Cover: Tarja Paalanen

Helsinki 2020

Finland

Author
Martin Vermeer

Name of the publication
Physical geodesy

Publisher School of Engineering

Unit Department of Built Environment

Series Aalto University publication series SCIENCE + TECHNOLOGY 2/2020

Field of research Geodesy

Language English

Abstract

Physical geodesy studies the large-scale figure and gravity field of the Earth, which are closely related. Our understanding of the gravity field is based on Newton's theory of gravitation. We present field theory, with partial differential equations describing the behaviour of the field throughout space. Techniques for solving these equations using boundary conditions on the Earth's surface are explained. A central concept is the geopotential.

The figure of the Earth is approximated by an ellipsoid of revolution, after which the precise figure is described by small deviations from this ellipsoid. Vertical reference systems are discussed in this context. Extending the approach to the Earth's gravity field yields small difference quantities, such as the disturbing potential and gravity anomalies.

Approaches to modelling the gravity field explained are spectral development of the field using spherical harmonics, the Stokes equation, numerical techniques based on the Fast Fourier Transform, the remove-restore technique, and least-squares collocation. Gravity measurement techniques are discussed, as are the multiple links with geophysics, such as terrain effects, isostasy, mean sea level and the sea level equation, and the tides.

Keywords figure of the Earth, gravity field, geopotential, reference ellipsoid, normal field, disturbing potential, gravity anomaly, geoid, height system, spherical harmonics, Stokes equation, remove-restore, least-squares collocation, gravimetry, isostasy, mean sea level, tides

ISBN (pdf) 978-952-60-8940-9

ISSN (PDF) 1799-490X

Location of publisher Helsinki **Location of printing** Helsinki **Year** 2020

Pages 524 **urn** http://urn.fi/URN:ISBN:978-952-60-8940-9

Preface

This book aims to present an overview of the current state of the study of the Earth's gravity field and those parts of geophysics closely related to it, especially geodynamics, the study of the changing Earth. It has grown out of over two decades of teaching at Helsinki's two universities: the Helsinki University of Technology — today absorbed into Aalto University — and the University of Helsinki. As such, it presents a somewhat Fennoscandian perspective on a very global subject. In addition, the author's own research on gravimetric geoid determination helped shape the presentation. While there exist excellent textbooks on all the different parts of what is presented here, I may still hope that this text will find a niche to fill.

Helsinki, 9th December, 2020,

Martin Vermeer

Second edition

A second, extensively corrected and improved edition was published on 25th September, 2022. No substantive content was added. The original

text was archived.

Acknowledgements

Thanks are due to the many students and colleagues, both in academia and at the Finnish Geodetic Institute, who have given useful comments and corrections over the course of many years of lecturing both at the University of Helsinki and at the Helsinki University of Technology, today Aalto University.

Special thanks are due to the foreign students at Aalto University, who forced me during recent years to provide an English version of this text. The translation work also prompted a basic revision of the Finnish text, which was long overdue as parts were written in the 1990s before the author had had the benefit of pedagogical training. Thanks are thus also due to Aalto University's pedagogical training programme.

Olivier Francis is gratefully acknowledged for contributing figure 11.8.

Several map images were drawn using the Generic Mapping Tools (Wessel et al., 2013).

The English language was competently checked by the Finnish Translation Agency Aakkosto Oy. Tarja Paalanen designed the cover. Laura Mure and Matti Yrjölä helped with the practicalities of publishing.

Contents

Chapters

List of Tables

List of Figures

Acronyms

ABCDEFGHIJKLMNOPRSTUW

A

ACS Advanced Camera for Surveys, instrument on the Hubble Space Telescope 2

AGU American Geophysical Union 402

B

BGI *Bureau Gravimétrique International*, International Gravity Bureau 126, 136, 154, 401, 402

BVP boundary-value problem 34

C

CHAMP Challenging Minisatellite Payload for Geophysical Research and Applications 77, 249, 377

D

DMA Defense Mapping Agency (USA) 76

DORIS Doppler Orbitography and Radiopositioning Integrated by Satellite, a French satellite positioning system 352, 353

DTM digital terrain model 135

E

EGM2008 Earth Gravity Model 2008 63, 77, 120, 126, 136

EGM96 Earth Gravity Model 1996 76–78

EGU European Geosciences Union 402
ENSO El Niño Southern Oscillation 329, 354
ESA European Space Agency 2, 341, 352, 353, 379

F

FAS member of the French Academy of Sciences (*Académie des sciences*) 433, 434
FFT Fast Fourier Transform 15, 153, 198, 214, 219, 234, 236–239, 241–247, 249, 250, 252, 253, 288, 291, 332, 439, 440
FGI Finnish Geospatial Research Institute, formerly Finnish Geodetic Institute 404
FIN2000 geoid model (Finland) 247, 248, 332
FIN2005N00 geoid model (Finland) 247, 403
FRAS Fellow of the Royal Astronomical Society 300
FRS Fellow of the Royal Society (of London) 4, 17, 226, 299, 300, 390, 391, 433, 434
FRSE Fellow of the Royal Society of Edinburgh 17, 390, 419, 434

G

GDR geophysical data record 356, 373
GFZ *Geoforschungszentrum* (Potsdam, Germany), German Research Centre for Geosciences 377
GIA glacial isostatic adjustment 333, 346
GNSS Global Navigation Satellite Systems, which comprise, besides American GPS, the satellite positioning systems of other countries, such as the Russian GLONASS and the European Galileo 116, 130, 187, 246, 247, 320, 321, 332, 333, 337, 352, 356, 372, 376, 380, 393, 394, 398, 403
GOCE Geopotential and Steady-state Ocean Circulation Explorer 77, 249, 288, 322, 339, 341, 376, 379–382
GPS Global Positioning System 70, 101, 173, 319, 325, 352, 377, 378
GRACE Gravity Recovery And Climate Experiment 77, 249, 377–380
GRAVSOFT Geopotential determination software, mainly developed in Denmark 247
GRS80 Geodetic Reference System 1980 6, 101, 104, 109, 112, 119, 247, 318, 332, 333
GWR20 superconducting gravimeter built by GWR Instruments 315

H

HUT Helsinki University of Technology, today part of Aalto University 404

I

IAG International Association of Geodesy 247, 315, 401, 402
IB inverted barometer 329
ICET International Center for Earth Tides 402
ICGEM International Center for Global Earth Models 402
IDEMS International Digital Elevation Model Service 402
IGeC International Geoid Commission (obsolete) 402
IGeS International Geoid Service (obsolete) 402
IGETS International Geodynamics and Earth Tide Service 315
IGFS International Gravity Field Service 401, 402
IMGC-02 transportable absolute rise-and-fall gravimeter, built by the *Istituto di Metrologia «G. Colonnetti»*, formerly the *Istituto Nazionale di Ricerca Metrologica* in Torino, Italy 312
IOC Initial Operating / Operational Capability 325
ISG International Service for the Geoid 401, 402
IUGG International Union of Geodesy and Geophysics 401

J

J_2 The dynamic flattening ("gravitational flattening") of the Earth 77, 101, 104, 367, 368, 382
Jason American-French-European radar altimetry satellite series, successors of TOPEX/Poseidon 352–354, 369, 382
JHU Johns Hopkins University 2

K

KKJ National Grid Co-ordinate System (Finland) 124

L

LAGEOS LAser GEOdynamic Satellite 366
Lego™ "*Leg Godt*", Engl. "*Play Well*", Danish toy brand 26
LLR Lunar laser ranging 308
LSC least-squares collocation 153

M

Mf Moon, fortnightly tide 390

N

N2000 height system (Finland) 165, 168, 247, 331
N60 height system (Finland) 164, 165, 247, 331, 332
NAP *Normaal Amsterdams Peil*, height system (Western Europe) 165
NASA National Aeronautics and Space Administration (USA) 2
NAVD88 North American Vertical Datum 1988 165
NC normal correction 185
NGA National Geospatial-Intelligence Agency (USA, formerly NIMA) 76, 401
NIMA National Imagery and Mapping Agency (USA, formerly DMA) 76
NKG *Nordiska Kommissionen för Geodesi*, Nordic Geodetic Commission 403
NKG2004 geoid model (Nordic area) 403
NKG2015 geoid model (Nordic area) 403
NN height system (Finland) 165
NOAA National Oceanic and Atmospheric Administration (USA) 341

O

OC orthometric correction 183, 184
OSU Ohio State University, Columbus, Ohio, USA 76

P

ppb parts per billion 120
ppm parts per million 120
PRARE Precise Range And Range-Rate Equipment, not operational 352
PRS President of the Royal Society 3, 122, 146, 419

R

RTM residual terrain modelling 215, 217, 218, 229

S

SI *Système international d'unités*, International System of Units 11, 37, 101, 120, 126
SK-42 Reference system of the Soviet Union, also known as the Krasovsky 1940 reference ellipsoid 102

SLR satellite laser ranging 352, 366
SRAL Synthetic Aperture Radar Altimeter 353
Ssa Sun, semi-annual tide 390
SST satellite-to-satellite tracking 379
STScI Space Telescope Science Institute 2
SWH significant wave height 355, 356

T

TC terrain correction 134, 135, 137–140, 250–252
TOPEX/Poseidon Topography Experiment / Poseidon, American-French radar altimetry satellite xix, 331, 352–354, 369, 382

U

UCO University of California Observatories 2

W

WGS84 World Geodetic System 1984 70, 101

Fundamentals of the theory of gravitation

1

1.1 General

In this chapter we present the foundations of Newton's theory of gravitation. Intuitively, the theory of gravitation is easiest to understand as "action at a distance," Latin *actio ad distans*, where the force between two masses is proportional to the masses themselves, and inversely proportional to the square of the distance between them. This is the form of Newton's general law of gravitation familiar to all.

There exists an alternative but equivalent presentation, *field theory*, which portrays gravitation as a phenomenon propagating through space, a *field*. The propagation is expressed in *field equations*. The field approach is not quite as intuitive, but is a powerful theoretical tool.[1] [1]

In this chapter we acquaint ourselves with the central concept of field theory, the *gravitational potential*. We also get to know the potential fields of the theoretically interesting single and double *mass-density layers*. Practical and theoretical applications of these include isostasy and

[1] There is also a philosophical side to this. To many, for example to Leibniz, the idea of a force that jumps from object to object was an abomination. Many tried to explain gravitation — and also electromagnetism, etc. — by a "world aether". It was not until the advent of relativity theory that the understanding gained ground that a physical theory does not have to satisfy our preconceptions about what constitutes a "sensible explanation" — as long as it presents the physical phenomena correctly.

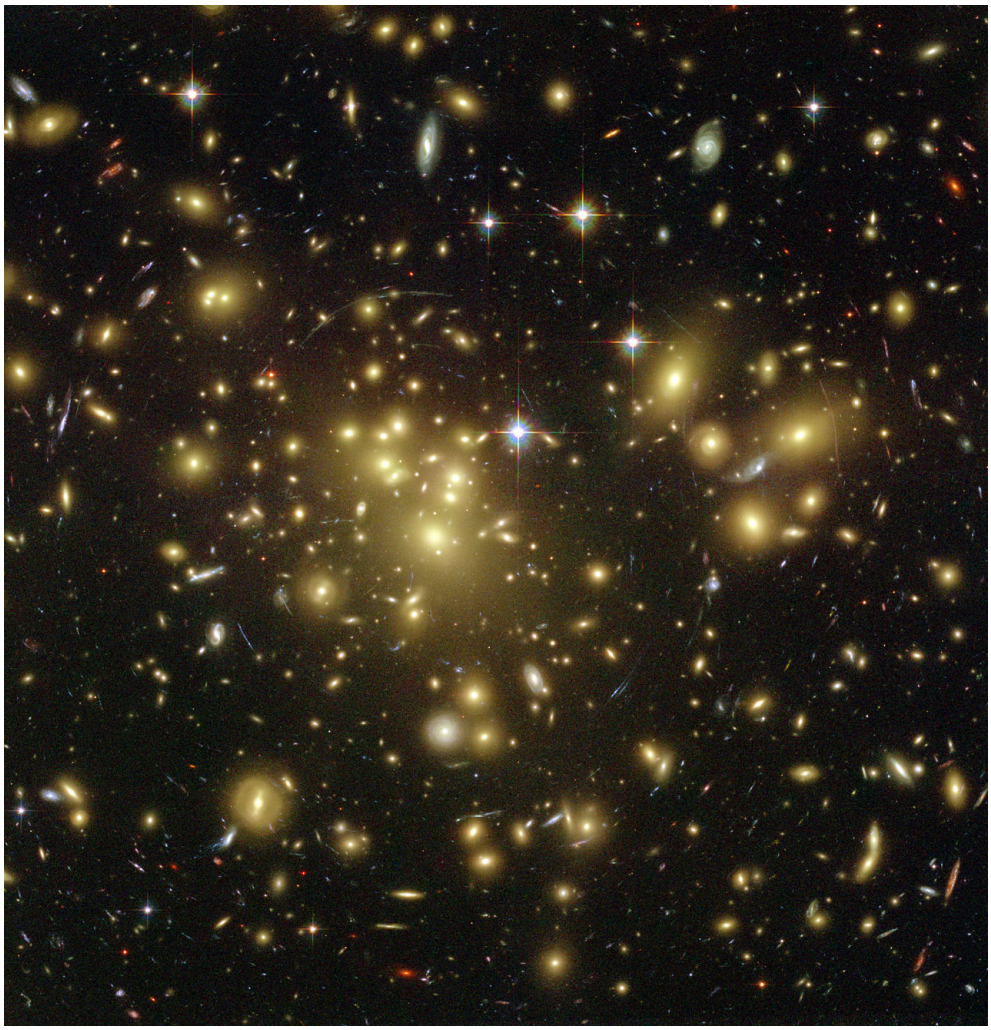


FIGURE 1.1. Gravitation is universal. A gravitational lens imaged by the Hubble Space Telescope, the Abell 1689 cluster of galaxies at a distance of 2.2 billion light years. Benitez et al. (2003).
Credits: NASA, N. Benitez (JHU), T. Broadhurst (The Hebrew University), H. Ford (JHU), M. Clampin (STScI), G. Hartig (STScI), G. Illingworth (UCO / Lick Observatory), the ACS Science Team and the ESA.

Helmert condensation, both of which we will study in later chapters.

We will learn about important *integral theorems* like the theorems

of Gauss and Green, with the aid of which we may infer the whole potential field in space from field values given only on a certain surface. Other similar examples are the Chasles theorem and the solution to Dirichlet's problem.

In chapters 2 and 3, we apply these fundamentals of potential theory to derive a spectral representation of the Earth's gravitational field, a *spherical-harmonic expansion*.

pallofunktio-kehitelmä

Here, at the beginning of the text, we derive an extensive set of mathematical equations, such as well-known integral equations. This is unfortunately necessary preliminary work. These equations, however, are not an end in themselves and are not worth committing to memory. Try rather to understand their logic, and how these various results have been arrived at historically. Try as well to acquire an intuition, a fingertip feeling, about the nature of this theory.

1.2 Gravitation between two masses

We start the investigation of the Earth's gravity field suitably with Isaac Newton's[2] general law of gravitation: 2

$$F = G\frac{m_1 m_2}{\ell^2}. \tag{1.1}$$

F is the attractional force between bodies 1 and 2, m_1 and m_2 are the masses of the bodies, and ℓ is the distance between them. We assume the masses to be points. The constant G, the *universal gravitational constant*, has the value

$$G = 6.674 \cdot 10^{-11}\ \mathrm{m^3/kg\,s^2}.$$

The value of G was determined for the first time by Henry Cavendish[3] 3

[2] Sir Isaac Newton PRS (1642–1727) was an English universal genius who derived the mathematical underpinning of astronomy, and much of geophysics, in his major work, *Philosophiæ Naturalis Principia Mathematica* (Mathematical Foundations of Physics).

torsiovaaka, kiertoheiluri

using a sensitive torsion balance (Cavendish, 1798).

Let us call the small body or test mass, for example a satellite, m, and the large mass, the planet or the Sun, M. Then, $m_1 = M$ may be called the *attracting* mass, and $m_2 = m$ the *attracted* mass, and we obtain

$$F = G\frac{mM}{\ell^2}.$$

According to Newton's law of motion

$$F = ma,$$

in which a is the gravitational acceleration of the body m. From this follows

$$a = G\frac{M}{\ell^2}.$$

From this equation, the quantity $m = m_2$ has vanished. This is
4 the famous observation by Galileo that all bodies fall equally fast,[4] irrespective of their mass. This is also known as Einstein's principle of
5 equivalence.[5]

Both the force F and the acceleration a have the same direction as the line connecting the bodies. For this reason one often writes equation 1.1 as a *vector equation*, which is more expressive:

$$\mathbf{a} = -GM\frac{\mathbf{r} - \mathbf{R}}{\ell^3}, \tag{1.2}$$

in which the three-dimensional vectors of place of both the attracting
6 and attracted masses are defined as follows in rectangular co-ordinates:[6]

[3] Henry Cavendish FRS (1731–1810) was a British natural scientist from a wealthy, noble background. He did also pioneering work in chemistry. He was extremely shy, and the renowned neurologist Oliver Sacks retrodiagnosed him as living with Asperger syndrome (Sacks, 2001).

[4] At least in a vacuum. The Apollo astronauts showed impressively how on the Moon a feather and a hammer fall equally fast! YouTube, Hammer vs. Feather.

[5] Albert Einstein (1879–1955) was a theoretical physicist of Jewish German descent, who created both the special and general theories of relativity, applying the latter to cosmology, and did fundamental work in quantum theory, a theory which he however never fully accepted.

$$\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}, \qquad \mathbf{R} = X\mathbf{i} + Y\mathbf{j} + Z\mathbf{k},$$

where the triad of unit vectors $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ is an *orthonormal basis*[7] in Euclidean space and

ortonormaali kanta
7

$$\ell = \|\mathbf{r} - \mathbf{R}\| = \sqrt{(x - X)^2 + (y - Y)^2 + (z - Z)^2} \tag{1.3}$$

is the distance between the masses computed by the Pythagoras theorem.

Vector equation 1.2 contains a minus sign, which tells us that the direction of the force is opposite to that of the vector $\mathbf{r} - \mathbf{R}$. This vector is the location of the attracted mass m reckoned from the location of the attracting mass M. In other words, this tells us that we are dealing with an *attraction* and not a repulsion.

1.3 The potential of a point mass

The gravitational field is a special field: if it is stationary, that is, not time-dependent, it is *conservative*. This means that a body moving inside the field along a closed path will, at the end of the journey, not have lost or gained energy. Thanks to this, one may affix to each point in the field a "label" onto which is written the amount of energy gained or lost by a unit or test mass when travelling from an agreed-upon *starting point* to the point under discussion. The value written on the label is called the *potential*.

Note that the choice of starting point is *arbitrary*. We shall return to this significant matter.

The *potential function* defined in this way for a point-like body M is

$$V = \frac{GM}{\ell}, \qquad \ell = \|\mathbf{r} - \mathbf{R}\|. \tag{1.4}$$

[6] As vector notation, one uses either $\vec{v}$ (an arrow above the letter) or $\boldsymbol{v}$ (bolding the letter). Here we use the bold notation where possible.

[7] This means that $\|\mathbf{i}\| = \|\mathbf{j}\| = \|\mathbf{k}\| = 1$ and $\langle \mathbf{i} \cdot \mathbf{j} \rangle = \langle \mathbf{i} \cdot \mathbf{k} \rangle = \langle \mathbf{j} \cdot \mathbf{k} \rangle = 0$, where the norm of vector $\mathbf{a}$ is defined as $\|\mathbf{a}\| \stackrel{\text{def}}{=} \sqrt{\langle \mathbf{a} \cdot \mathbf{a} \rangle}$, and $\langle \mathbf{a} \cdot \mathbf{b} \rangle$ is the *inner* or *scalar product* of vectors $\mathbf{a}$ and $\mathbf{b}$ defined in the space.

The constant GM has in the case of the Earth (according to the GRS80, conventionally) the value

$$GM_\oplus = 3.986\,005 \cdot 10^{14}\ \mathrm{m^3/s^2}.$$

The currently best value based on measurement is slightly more precise:

$$GM_\oplus = 3.986\,004\,418\,(8) \cdot 10^{14}\ \mathrm{m^3/s^2}.$$

The number in parentheses (8) represents the uncertainty in units of the last given decimal. The relative uncertainty is thus $2 : 10^9$.

1.4 Potential of a spherical shell

We may write, based on equation 1.4, the potential of an extended body M into the following form:

$$V(\mathbf{r}) = G\int_M \frac{dm(\mathbf{R})}{\ell} = G\int_M \frac{dm(\mathbf{R})}{\|\mathbf{r}-\mathbf{R}\|}. \tag{1.5}$$

This is an integral over the mass elements dm of the body, where every such mass element is located at its own place $\mathbf{R}$. The potential V is evaluated at location $\mathbf{r}$, and the distance $\ell = \|\mathbf{r}-\mathbf{R}\|$.

We now derive the equation for the potential of a thin spherical shell, see figure 1.2, where we have placed the centre of the sphere at the origin O.

Because the circumference of a narrow ring, width $b \cdot d\theta$, is $2\pi b \sin\theta$, its surface area is

$$(2\pi b \sin\theta)\,(b \cdot d\theta)\,.$$

Let the thickness of the shell be p (small) and its density ρ. We obtain for the total mass of the ring

$$2\pi p \rho b^2 \sin\theta\, d\theta.$$

Because every point of the ring is at the same distance ℓ from point P, we may write for the potential of the ring at point P:

$$V_P = \frac{2\pi G p \rho b^2 \sin\theta\, d\theta}{\ell}.$$

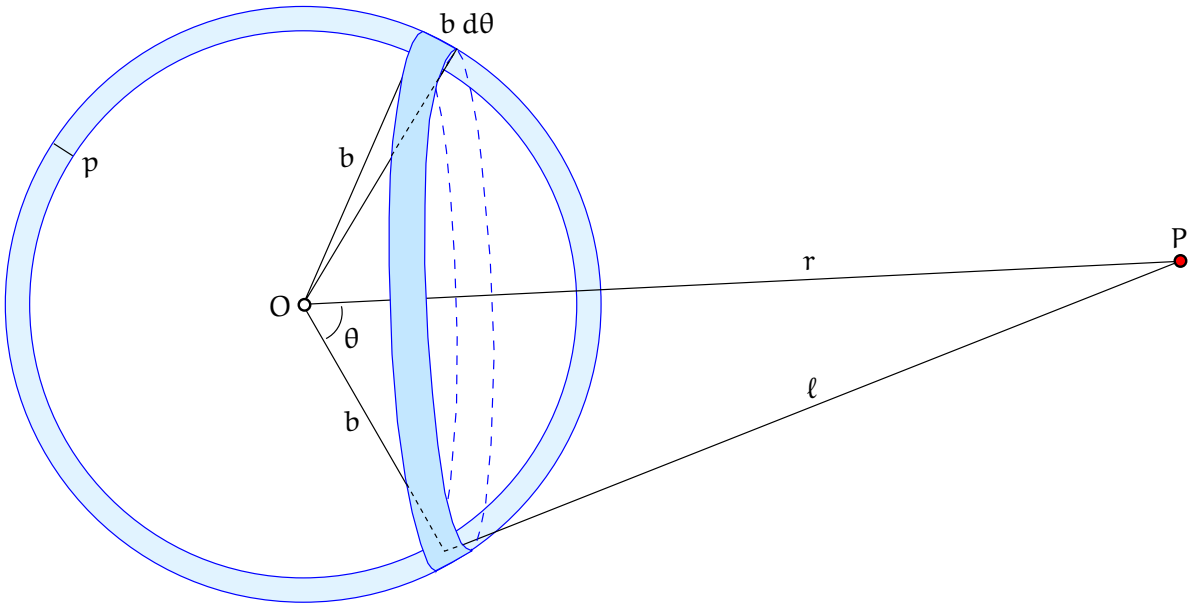


FIGURE 1.2. A thin spherical shell consists of rings.

With the cosine rule,

$$\ell^2 = r^2 + b^2 - 2rb\cos\theta, \tag{1.6}$$

we obtain, using equation 1.5, for the potential of the whole shell

$$V_P = 2\pi G\rho p b^2 \int \frac{\sin\theta\, d\theta}{\sqrt{r^2 + b^2 - 2rb\cos\theta}}.$$

In order to evaluate this integral, we must replace the integration variable θ by ℓ. Differentiating equation 1.6 yields

$$\ell\, d\ell = rb\sin\theta\, d\theta,$$

and remembering that $\ell = \sqrt{r^2 + b^2 - 2rb\cos\theta}$ we obtain

$$V_P = 2\pi G\rho p b^2 \int_{\ell_1}^{\ell_2} \frac{d\ell}{rb}.$$

In the case that point P is outside the shell, the integration bounds of ℓ are $\ell_1 = r - b$ and $\ell_2 = r + b$, and we obtain for the potential of point P

$$V_P = 2\pi G\rho p b^2 \left[\frac{\ell}{rb}\right]_{\ell=r-b}^{\ell=r+b} = \frac{4\pi G\rho p b^2}{r}.$$

Because the mass of the whole shell is $M_b = 4\pi b^2 \rho p$, it follows that the potential of the shell is *the same as that of an equal-sized mass in its centre* O:

$$V_P = \frac{GM_b}{r},$$

where r is the distance of computation point P from the centre of the sphere O. We see that this is identical to equation 1.4.

In the same way, the attraction, or rather, *acceleration*, caused by the spherical shell is[8]

$$\mathbf{a}_P = \nabla V|_P = -4\pi G\rho p b^2 \frac{\mathbf{r}_P - \mathbf{r}_O}{r^3} = -GM_b \frac{\mathbf{r}_P - \mathbf{r}_O}{r^3},$$

in which $r = \|\mathbf{r}_P - \mathbf{r}_O\|$. This result is identical to the acceleration caused by an equal-sized point mass located in point O, equation 1.2.

In the case that point P is inside the shell, $\ell_1 = b - r$ and $\ell_2 = b + r$ and the above integral changes to the following:

$$V_P = 2\pi G\rho p b^2 \left[\frac{\ell}{rb}\right]_{\ell=b-r}^{\ell=b+r} = 4\pi G\rho p b.$$

As we see, this is a *constant* and not dependent upon the location of point P. Therefore $\nabla V_P = 0$ and the attraction, being the gradient of the potential, vanishes.

The end result is that the attraction of a spherical shell is, outside the shell,

$$a = \|\mathbf{a}\| = \frac{GM}{r^2},$$

where M is the total mass of the shell and $r = \|\mathbf{r}_P - \mathbf{r}_O\|$ the distance of the observation point from the shell's centre. The attraction vanishes inside the shell.

In figure 1.3 we have drawn the curves of potential and attraction — or more precisely acceleration, attractive force per unit of mass. If a body consists of many concentric spherical shells (like rather precisely

[8] Here, the ∇ (nabla) operator is used, to be explained in section 1.5.

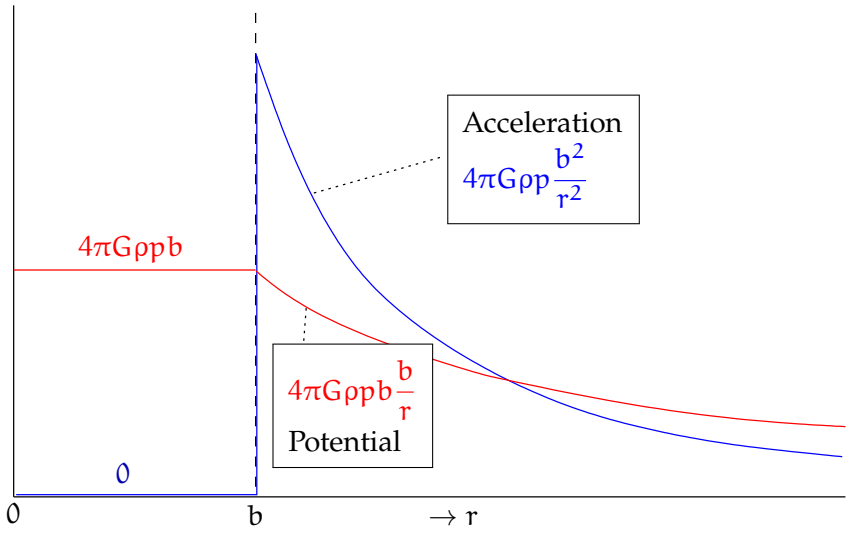


FIGURE 1.3. Dependence of potential and attraction on distance r from the centre of a spherical shell.

the Earth and many celestial bodies), then inside the body only those layers of mass "interior" to — meaning closer to the centre than — the observation point participate in causing attraction, and this attraction is the same as it would be if the mass of all these layers was concentrated in the centre of the body. This case, where the distribution of mass density inside a body only depends on the distance from its centre, is called an *isotropic* density distribution.

1.5 Computing the attraction from the potential

As we argued earlier, the potential V is a *path integral*. Conversely we can compute, from the potential, the components of the gravitational acceleration vector by *differentiating* $V(x, y, z)$ with respect to *place*, by applying the *gradient operator*, which is a vector operator:

$$\mathbf{a} = \nabla V = \operatorname{grad} V = \frac{\partial V}{\partial x}\mathbf{i} + \frac{\partial V}{\partial y}\mathbf{j} + \frac{\partial V}{\partial z}\mathbf{k}. \tag{1.7}$$

Here, the symbol ∇ (nabla), is a frequently used *partial differential operator*

$$\nabla = \mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z}.$$

kanta As before, $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ is a *basis* of mutually orthogonal unit vectors in Euclidean space parallel to the (x, y, z) axes.

Let us try this differentiation in the case of the potential field of the point mass M. Substitute the above equations for potential V and
9 distance ℓ, 1.3 and 1.4:[9]

$$\frac{\partial V}{\partial x} = \frac{\partial V}{\partial \ell}\frac{\partial \ell}{\partial x} = GM \cdot -\frac{1}{\ell^2} \cdot \frac{x - X}{\ell} = -GM\frac{x - X}{\ell^3}.$$

Similarly we compute the y and z components:

$$\frac{\partial V}{\partial y} = -GM\frac{y - Y}{\ell^3}, \quad \frac{\partial V}{\partial z} = -GM\frac{z - Z}{\ell^3}.$$

These are the components of the gravitational acceleration or attraction vector when the source of the field is one point mass M. So, in this concrete case, vector equation 1.7 given above applies:

$$\mathbf{a} = \operatorname{grad} V = \nabla V.$$

Remark In physical geodesy — unlike in physics — the potential is reckoned to be always positive if the attracting mass M is positive,

---

[9] From the equation

$$\ell = \sqrt{(x - X)^2 + (y - Y)^2 + (z - Z)^2} = \left((x - X)^2 + (y - Y)^2 + (z - Z)^2\right)^{1/2}$$

it follows with the chain rule that

$$\frac{\partial \ell}{\partial x} = \frac{\partial \left((x - X)^2 + (y - Y)^2 + (z - Z)^2\right)^{1/2}}{\partial\left((x - X)^2 + (y - Y)^2 + (z - Z)^2\right)} \cdot \frac{\partial (x - X)^2}{\partial x} =$$

$$= \tfrac{1}{2}\left((x - X)^2 + (y - Y)^2 + (z - Z)^2\right)^{-1/2} \cdot 2(x - X) = \frac{x - X}{\ell}.$$

as it is known to always be. However, the *potential energy* of body m inside the field V of mass M is negative! More precisely, the potential energy of body m is

$$E_{pot} = -Vm.$$

One calls the vector of gravitational acceleration more briefly the *gravitation* or *attraction* vector.

1.6 Potential of a solid body

In the following we study a *solid body*, the mass of which is distributed throughout space and thus not concentrated in a single point. For example, the mass distribution of the Earth in space may be described by a matter *density function* ρ:

$$\rho(x, y, z) = \frac{dm(x, y, z)}{dV(x, y, z)},$$

in which dm is a mass element and dV is the corresponding element of spatial volume. The dimension of ρ is density, its unit in the SI system, $^{kg}/_{m^3}$.

Because the gravitational acceleration 1.7 is a linear expression in the potential V, and both the force and acceleration vectors may be summed linearly, it follows that the total potential of the body can also be obtained by summing together the potentials of all its parts. For example, the potential of a set of n mass points is

$$V(\mathbf{r}) = G \sum_{i=1}^{n} \frac{m_i}{\ell_i} = G \sum_{i=1}^{n} \frac{m_i(\mathbf{R}_i)}{\|\mathbf{r} - \mathbf{R}_i\|},$$

from which we obtain the gravitational acceleration by simply using the gradient theorem 1.7.

The potential of a solid body is obtained similarly by replacing the sum by an integral, in the following way:[10]

[10]

$$V = G \iiint_{\text{body}} \frac{dm}{\ell} = G \iiint_{\text{body}} \frac{\rho}{\ell} d\mathcal{V}. \tag{1.8}$$

The symbol ρ inside the integral stands for matter density at the location of volume element $d\mathcal{V}$. The quantity $\ell = \|\mathbf{r} - \mathbf{R}\| = \sqrt{(x-X)^2 + (y-Y)^2 + (z-Z)^2}$ is the distance between the point at which the potential is being calculated, and the attracting mass element. More clearly:

$$V(x,y,z) = G \iiint_{\text{body}} \frac{\rho(X,Y,Z)}{\sqrt{(x-X)^2 + (y-Y)^2 + (z-Z)^2}} \, dX \, dY \, dZ.$$

As we showed already above for mass points, the first derivative with respect to place or *gradient* of the potential V of a solid body,

$$\operatorname{grad} V = \nabla V = \mathbf{a}, \tag{1.9}$$

is also the acceleration vector caused by the attraction of the body. This applies generally.

1.6.1 Behaviour at infinity

If a body is of finite extent — in other words, it lies completely within a sphere of size ϵ around the origin — and its density is also bounded everywhere, it follows that

$$\|\mathbf{r}\| \to \infty \quad \Longrightarrow \quad V(\mathbf{r}) \to 0,$$

because, according to the triangle inequality,

$$\ell = \|\mathbf{r} - \mathbf{R}\| \geqslant \|\mathbf{r}\| - \|\mathbf{R}\| \geqslant \|\mathbf{r}\| - \epsilon$$

and thus

$$\|\mathbf{r}\| \to \infty \quad \Longrightarrow \quad 1/\ell \to 0.$$

[10] Unfortunately almost the same symbols V and $\mathcal{V}$ are used here for potential and volume, respectively.

For the acceleration of gravitation, the same applies for all three components, and thus also for the length of this vectorial quantity:

$$\|\mathbf{r}\| \to \infty \quad \Longrightarrow \quad \|\nabla V\| \to 0.$$

This result can still be sharpened: if $\|\mathbf{r}\| \to \infty$, then, again by the triangle inequality,

$$\ell = \|\mathbf{r} - \mathbf{R}\| \leqslant \|\mathbf{r}\| + \|\mathbf{R}\| \leqslant \|\mathbf{r}\| + \epsilon,$$

and thus

$$\frac{1}{\|\mathbf{r}\| + \epsilon} \leqslant \frac{1}{\ell} \leqslant \frac{1}{\|\mathbf{r}\| - \epsilon} \implies \frac{1}{\|\mathbf{r}\|} \frac{1}{1 + \epsilon/\|\mathbf{r}\|} \leqslant \frac{1}{\ell} \leqslant \frac{1}{\|\mathbf{r}\|} \frac{1}{1 - \epsilon/\|\mathbf{r}\|}.$$

It is seen, again with the notation $r = \|\mathbf{r}\|$, that

$$r \to \infty \quad \Longrightarrow \quad 1/\ell \to 1/r\,.$$

When we substitute this into integral 1.8, it follows that for large distances $r \to \infty$:

$$V = G \iiint_{\text{body}} \frac{\rho}{\ell}\, d\mathcal{V} \approx \frac{G}{r} \iiint_{\text{body}} \rho\, d\mathcal{V} = \frac{GM}{r},$$

in which M, the density integrated over the volume of the body, is precisely its *total mass*. From this we see that, at great distance, the field of a finite-sized body M is *nearly identical* with the field generated by a point mass the mass of which is equal to the *total mass* of the body. This important observation was already made by Newton. Thanks to this phenomenon, we may treat, in the celestial mechanics of the solar system, the Sun and planets[11] as mass points, although we know that they are not.

[11]

[11] The only important exception is formed by the forces between a planet and its moons, due both to the flattening of the planet and tidal effects.

1.7 Example: The potential of a line of mass

The potential of a vertical line of mass having a linear mass density of unity is

$$V(x,y,z) = G\int_0^H \frac{1}{\sqrt{(X-x)^2+(Y-y)^2+(Z-z)^2}}\,dZ, \tag{1.10}$$

in which (X,Y) is the location in the plane of the mass line, (x,y,z) is the location of the point at which the potential is being evaluated, and the mass line extends from sea level $Z=0$ to height $Z=H$.

Firstly we write $\Delta x = X-x$, $\Delta y = Y-y$, and $\Delta z = Z-z$, and the potential becomes

$$V(\Delta x,\Delta y,\Delta z) = G\int_{-z}^{H-z} \frac{1}{\sqrt{\Delta x^2+\Delta y^2+\Delta z^2}}\,d(\Delta z).$$

The indefinite integral is

$$\ln\left(\Delta z+\sqrt{\Delta x^2+\Delta y^2+\Delta z^2}\right)$$

and substituting the integration bounds yields

$$V = G\ln\frac{H-z+\sqrt{\Delta x^2+\Delta y^2+(H-z)^2}}{-z+\sqrt{\Delta x^2+\Delta y^2+z^2}}.$$

Expand this into a Taylor series in H around the point $H=0$: the first derivative of equation 1.10 is

$$\frac{\partial V}{\partial H} = \frac{G}{\sqrt{(X-x)^2+(Y-y)^2+(H-z)^2}} = \frac{G}{\ell}$$

in which $\ell(H) = \sqrt{(X-x)^2+(Y-y)^2+(H-z)^2}$. The second derivative is, using the chain rule,

$$\frac{\partial^2 V}{\partial H^2} = \frac{\partial}{\partial H}\left(\frac{G}{\ell}\right) = G\cdot\frac{\partial \ell^{-1}}{\partial \ell}\cdot\frac{\partial \ell}{\partial H} = \\ = G\cdot -\ell^{-2}\cdot\tfrac{1}{2}\ell^{-1}\cdot 2\,(H-z) = -G\frac{H-z}{\ell^3}.$$

The third derivative, obtained in the same way:

$$\frac{\partial^3 V}{\partial H^3} = -G\frac{\partial}{\partial H}\left(\frac{H-z}{\ell^3}\right) = G\left(\frac{3\,(H-z)^2}{\ell^5} - \frac{1}{\ell^3}\right) = \\ = G\frac{3\,(H-z)^2 - \ell^2}{\ell^5},$$

and so on. The Taylor expansion is

$$V = \underbrace{V|_{H=0}}_{0} + \overbrace{G\frac{1}{\ell_0}}^{\partial_H V|_{H=0}} H + \frac{1}{2}\overbrace{G\frac{z}{\ell_0^3}}^{\partial_H^2 V|_{H=0}} H^2 + \frac{1}{6}\overbrace{G\frac{3z^2-\ell_0^2}{\ell_0^5}}^{\partial_H^3 V|_{H=0}} H^3 + \cdots, \qquad (1.11)$$

in which $\ell_0 = \sqrt{(X-x)^2 + (Y-y)^2 + z^2}$, so the values of the derivatives used in this expansion are obtained by substituting $H = 0$.

Question How can we exploit this result for computing the gravitational potential of a complete, realistic terrain or topography?

Answer In this expansion, the coefficients $1/\ell_0$, $\frac{1}{2}\,z/\ell_0^3$, …, like ℓ_0, depend only on the *differences* $\Delta x = X - x$ and $\Delta y = Y - y$ between the co-ordinates of the location of the mass line (X, Y) and those of the evaluation location (x, y) — and of the height z of the evaluation point. If the topography is given in the form of a grid, we may evaluate the above expansion 1.11 term-wise for the given z value and for all possible value pairs $(\Delta x, \Delta y)$.

Then, if the grid is, say, $N \times N$ in size, we need only N^2 operations to calculate every coefficient. The brute-force evaluation of the Taylor expansion itself *for the whole topography*, for every point of the terrain grid and every point of the evaluation grid, requires after that $N^2 \cdot N^2 = N^4$ operations, but those are much simpler: the coefficients themselves have been precomputed. And brute force is not even the best approach: as we shall see, the above *convolution* can be computed much faster using the Fast Fourier Transform (FFT).

We shall return to this subject more extensively in connection with terrain-effect evaluation, in sections 6.3 and 9.7.

1.8 The equations of Laplace and Poisson

The *second derivative* with respect to the place of the geopotential, the first derivative with respect to the place of the gravitational acceleration vector called its *divergence*, is also of geophysical interest. We may write:

$$\operatorname{div} \mathbf{a} \stackrel{\text{def}}{=} \langle \nabla \cdot \mathbf{a} \rangle = \langle \nabla \cdot (\nabla V) \rangle = \langle \nabla \cdot \nabla \rangle V = \\ = \Delta V = \frac{\partial^2}{\partial x^2} V + \frac{\partial^2}{\partial y^2} V + \frac{\partial^2}{\partial z^2} V, \quad (1.12)$$

in which

$$\Delta \stackrel{\text{def}}{=} \langle \nabla \cdot \nabla \rangle = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

is a well-known symbol called the *Laplace operator.*[12]

In equation 1.4 for the potential of a point mass we may show, by performing all partial derivations 1.12, that

$$\Delta V = 0, \quad (1.13)$$

the well-known *Laplace equation*. This equation applies outside a point mass, and more generally everywhere in empty space: all masses can in the limit be considered to consist of point-like mass elements. Or, in equation 1.8 we may directly differentiate inside the triple integral sign, exploiting the circumstance that it is permitted to interchange integration and partial differentiation, if both exist.

A potential field for which Laplace equation 1.13 applies, is called a *harmonic* field.

[12] Pierre-Simon marquis de Laplace (1749–1827) was a French universal genius who contributed to mathematics and the natural sciences. He is one of the 72 French scientists, engineers and mathematicians whose names were inscribed on the Eiffel Tower, Eiffel Tower, 72 names.

In the case where the mass density does not vanish everywhere, we have a different equation, with ρ the mass density:

$$\Delta V = -4\pi G\rho. \tag{1.14}$$

This equation is called the *Poisson equation*.[13]

[13]

The pair of equations

$$\operatorname{grad} V = \mathbf{a}, \qquad \operatorname{div} \mathbf{a} = -4\pi G\rho$$

are known as the *field equations* of the gravitational field. They play the same role as Maxwell's[14] field equations in electromagnetism. Unlike in Maxwell's equations, however, in the above there is no time coordinate. Because of this, it is not possible to derive an equation describing the propagation in space of gravitational waves, like the one for electromagnetic waves in Maxwell's theory.

[14]

We know today that these "Newton field equations" are only approximate, and that Einstein's general theory of relativity is a more precise theory. Nevertheless, in physical geodesy Newton's theory is generally precise enough and we shall use it exclusively.

1.9 Gauge invariance

An important property of the potential is that if we add a *constant* C to it, nothing related to gravitation that can actually be measured, changes. This is an example of what is called *gauge invariance*.

mitta-invarianssi

Gravitation itself is obtained by differentiation of the potential, an operation that eliminates the constant term. Therefore the definition

---

[13] Siméon Denis Poisson (1781–1840) was a French mathematician, physicist and geodesist, one of the 72 names inscribed on the Eiffel Tower, Eiffel Tower, 72 names.

[14] James Clerk Maxwell FRS FRSE (1831–1879) was a Scottish physicist, the discoverer of the field equations of electromagnetism. He found a wave-like solution to the equations, and, based on propagation speed, identified *light* as such.

of potential is ambiguous: all potential fields V obtained by different choices of C are equally valid.

Observations only give us potential *differences*, as spirit levellers know very well.

An often-chosen definition of potential is based on requiring that, if $r = \|\mathbf{r}\| \to \infty$, then also $V \to 0$, which makes physical sense and yields simple equations. However, in work on the Earth's surface, a more practical alternative may be $V = 0$ at the mean sea surface — although that is not free of problems, either.

For example, for the mass of the Earth $M_\oplus$ a physically sensible form of the potential is, in spherical approximation,

$$V = \frac{GM_\oplus}{r},$$

which vanishes at infinity $r \to \infty$, when again a geodetically sensible definition would be

$$V = \frac{GM_\oplus}{r} - \frac{GM_\oplus}{R},$$

in which $R = \|\mathbf{R}\|$ is the radius of the Earth sphere. The latter potential vanishes where $r = R$, on the surface of a spherical Earth, or "sea level". In the limit $r \to \infty$ the value of the potential is $-GM_\oplus/R$, not zero.

1.10 Single mass-density layer

If we apply to the surface S of a body a mass "coating", of mass-density value

$$\kappa = \frac{dm}{dS},$$

we obtain for the potential an integral equation otherwise similar to equation 1.8, but a *surface* integral:

$$V = G \iint_{\text{surface}} \frac{dm}{\ell} = G \iint_{\text{surface}} \frac{\kappa}{\ell}\, dS. \tag{1.15}$$

Here again ℓ is the distance between the point at which the potential is to be calculated and the moving mass element in integration dm —

or surface element dS. The dimension of the mass surface density κ is $^{kg}/_{m^2}$, different from the dimension of ordinary or volume mass density, which is $^{kg}/_{m^3}$.

This case is theoretically interesting, although physically unrealistic. The function V is everywhere continuous, also at the surface S, however its first derivatives with place are already discontinuous. The discontinuity appears in the direction perpendicular to the surface, in the *normal derivative*.

Let us look at the simple case where a sphere of radius R has been coated with a layer of constant surface density κ. By computing the above integral 1.15 we may prove — in a complicated way, see section 1.4 — that the exterior potential is the same as it would be if all of the mass of the coating were concentrated in the sphere's centre. Also in section 1.4 we proved that the potential interior to the sphere is constant.

Thus, the exterior attraction $(r > R)$, with r the distance of the point of computation from the centre of the sphere, is

$$a_{ext}(r) = G\frac{M}{r^2} = G\frac{\kappa \cdot 4\pi R^2}{r^2} = 4\pi G\kappa \left(\frac{R}{r}\right)^2.$$

The interior attraction $(r < R)$ is

$$a_{int}(r) = 0.$$

This means that on the surface of the sphere, $\ell = R$, the attraction is *discontinuous*:

$$a_{ext}(R) - a_{int}(R) = 4\pi G\kappa.$$

In this symmetric case we see that

$$a = \|\mathbf{a}\| = \frac{\partial V}{\partial n}, \tag{1.16}$$

in which the differentiation variable n represents the *normal direction*, the direction perpendicular to the surface S. If the surface S is an *equipotential surface* of the potential V, equation 1.16 applies generally. Then, the attraction vector — more precisely, the acceleration vector — is perpendicular to the surface S, and its magnitude is equal to that of the normal derivative of the potential.

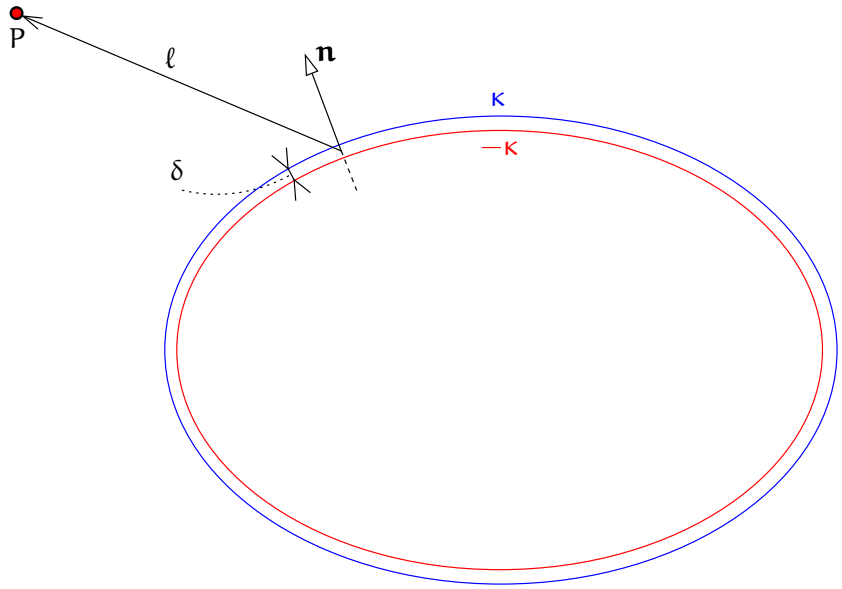


FIGURE 1.4. A double mass-density layer.

1.11 Double mass-density layer

A double mass-density layer may be interpreted as a *dipole-density layer*. The dipoles are oriented in the direction of the surface normal.

If the dipole consists of two "charges" m and $-m$ in locations $\mathbf{r}_1$ and $\mathbf{r}_2$, in such a way that the vectorial separation between them is $\Delta\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$, then the dipole *moment* is $\mathbf{d} = m\,\Delta\mathbf{r}$, a vectorial quantity. See figure 1.4.

Let the dipole surface density be

$$\mu = \frac{d\mathcal{D}}{dS},$$

in which $d\mathcal{D}$ is a "dipole-layer element". This layer may be seen as being made up of two single layers. If we have a positive layer at density κ and a negative layer at density $-\kappa$, and the distance between them is δ, we get for small values of δ an approximate correspondence:

$$\mu \approx \kappa\delta. \tag{1.17}$$

The combined potential of the two single mass-density layers computed as explained in the previous section, equation 1.15, is

$$V = G \iint_{\text{surface}} \kappa \left(\frac{1}{\ell_1} - \frac{1}{\ell_2} \right) dS.$$

The following relationship exists between the quantities ℓ_1, ℓ_2 and δ (Taylor expansion of the function $1/\ell$):

$$\frac{1}{\ell_1} = \frac{1}{\ell_2} + \delta \cdot \frac{\partial}{\partial n}\left(\frac{1}{\ell}\right) + \cdots,$$

in which $\frac{\partial}{\partial n}$ is again the derivative of the quantity in the normal direction of the surface.

Substitution into the equation yields

$$V \approx G \iint_{\text{surface}} \kappa\delta \frac{\partial}{\partial n}\left(\frac{1}{\ell}\right) dS = G \iint_{\text{surface}} \mu \frac{\partial}{\partial n}\left(\frac{1}{\ell}\right) dS. \qquad (1.18)$$

In the limit in which δ is arbitrarily small and κ correspondingly large, this equation, like equation 1.17, holds exactly.

One can easily show that the above potential is not even continuous. The discontinuity happens at the surface S. Let us look again, for the sake of simplicity, at a sphere of radius R, coated with a double layer of constant dipole density μ.

The exterior potential ($r > R$, r the distance from the centre of the sphere) is

$$V_{\text{ext}} = G\mu \iint_{\text{surface}} \frac{\partial}{\partial n}\left(\frac{1}{\ell}\right) dS = 0,$$

because the potential equals the difference between the potentials of two concentric spherical shells of the same mass.

The interior potential ($r < R$) is

$$V_{\text{int}} = G\mu \iint_{\text{surface}} \frac{\partial}{\partial n}\left(\frac{1}{\ell}\right) dS = G\mu \cdot 4\pi R^2 \left(-\frac{1}{R^2}\right) = -4\pi G\mu,$$

by computing the surface integral using the sphere's centre as the evaluation point, and using the earlier established circumstance that

inside a sphere covered by a *single* constant-density mass-density layer the potential is constant.

Now in the limit $r \to R$ the result is different for the exterior and interior potentials. The difference is

$$V_{ext}(R) - V_{int}(R) = 4\pi G\mu.$$

1.12 The Gauss divergence theorem

1.12.1 Presentation

[15] The Gauss divergence theorem,[15] famous from physics, looks in vector form like this:

$$\iiint_{\mathcal{V}} \operatorname{div} \mathbf{a}\, d\mathcal{V} = \iint_{\partial\mathcal{V}} \langle \mathbf{a} \cdot \mathbf{n} \rangle\, dS, \tag{1.19}$$

in which $\mathbf{n}$ is the exterior normal to surface S, now as a vector: the length of the vector is assumed $\|\mathbf{n}\| = 1$. $\partial\mathcal{V}$ is the total surface of body $\mathcal{V}$.

This theorem applies to all differentiable vector fields $\mathbf{a}$ and all "well-behaved" bodies $\mathcal{V}$ on whose surface $\partial\mathcal{V}$ a normal direction $\mathbf{n}$ exists everywhere. In other words, this is not a special property of the gravitational acceleration vector, though it also applies to this vector field.

1.12.2 Intuitive description

[16] Note that[16]

$$\operatorname{div} \mathbf{a} = \Delta V = -4\pi G\rho$$

is a *source function*. It stands for the amount, in the part of space inside surface $\partial\mathcal{V}$, of positive and negative "sources" and "sinks" of the

lähteet, nielut

[15] Johann Carl Friedrich Gauss (1777–1855) was a German mathematician and universal genius. "*Princeps mathematicorum*".

[16] Assuming that for the vector field $\mathbf{a}$ a potential V exists, see section A.4.

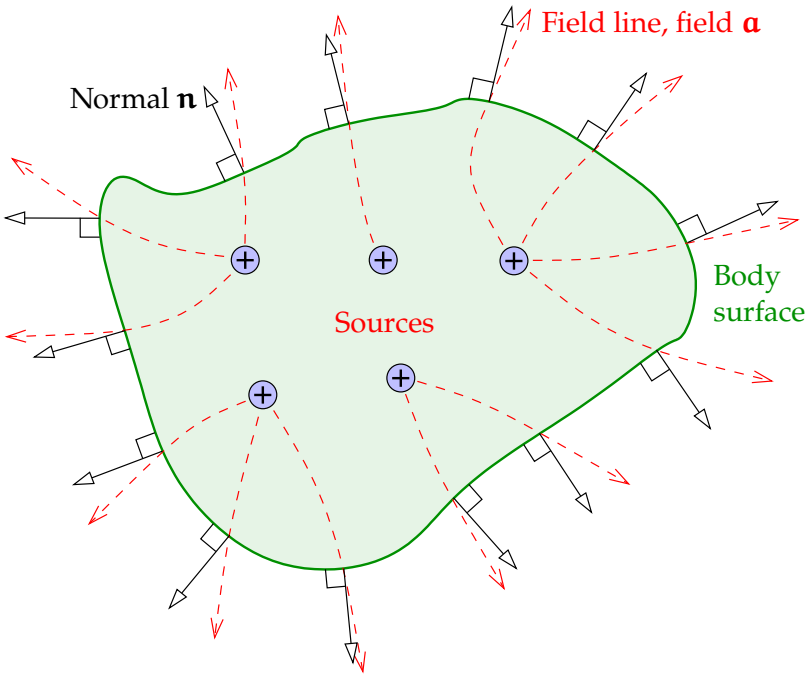


FIGURE 1.5. A graphical explanation of the Gauss divergence theorem. The concept of the *field line* was Michael Faraday's invention. The *flux* is the scalar product $\langle \mathbf{a} \cdot \mathbf{n} \rangle$.

gravitational field.

The situation is analogous to the flow pattern of a liquid: positive charges correspond to points where liquid is added to the flow, negative charges[17] correspond to "sinks" through which liquid disappears. The vector $\mathbf{a}$ is in this metaphor the velocity of flow; in the absence of "sources" and "sinks" it satisfies the condition $\operatorname{div} \mathbf{a} = 0$, which expresses the conservation and incompressibility of matter.

On the other hand, the function

$$\langle \mathbf{a} \cdot \mathbf{n} \rangle = \frac{\partial V}{\partial n}$$

is often called the *flux*, in other words, how much field stuff "flows out"

[17] But the "charges" for gravitation, the masses, are always positive.

— just like a liquid flow — from the space inside the surface $\partial\mathcal{V}$ through the surface to the outside.

The Gauss divergence theorem states the two amounts to be equal: it is in a way a *book-keeping statement* demanding that everything which is produced inside a surface — div $\mathbf{a}$ — also has to come out through the surface — $\langle \mathbf{a} \cdot \mathbf{n} \rangle$.

In figure 1.5 it is graphically explained that the sum of "sources" over the inner space of the body, $\sum(+++\cdots)$, has to be the same as the sum of "flux" $\sum(\uparrow\uparrow\uparrow\cdots)$ over the whole boundary surface delimiting this inner space.

1.12.3 *The potential version of the Gauss divergence theorem*

Let us write the Gauss divergence theorem a little differently, using the *potential* V instead of the gravitational vector:

$$\iiint_{\mathcal{V}} \Delta V \, d\mathcal{V} = \iint_{\partial\mathcal{V}} \frac{\partial V}{\partial n} \, dS, \tag{1.20}$$

in which we have made the above substitutions. We also here see the notation $\partial\mathcal{V}$ for the surface of the body $\mathcal{V}$. The presentational forms 1.20 and 1.19 are connected by equations 1.12 and 1.9, connecting V and $\mathbf{a}$.

1.12.4 *Example 1: A little box*

Let us look at a little rectangular box with sides Δx, Δy, and Δz. The box is so little that the field $\mathbf{a}(x, y, z)$ inside it is an almost linear function of place. Let us write the vector $\mathbf{a}$ into components:

$$\mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}.$$

Now the volume integral

$$\iiint_{\mathcal{V}} \operatorname{div} \mathbf{a} \, d\mathcal{V} \approx \left(\frac{\partial a_1}{\partial x} + \frac{\partial a_2}{\partial y} + \frac{\partial a_3}{\partial z} \right) \Delta x \, \Delta y \, \Delta z \tag{1.21}$$

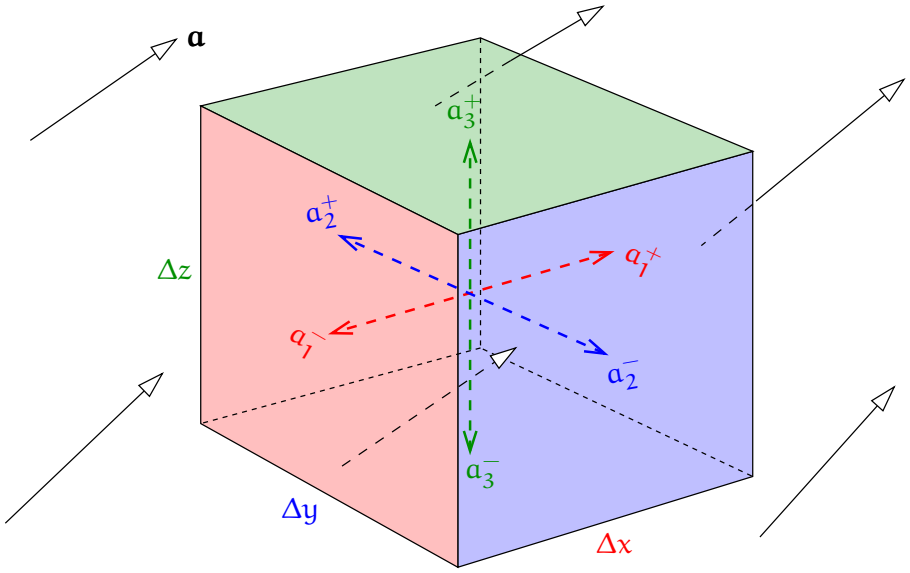


FIGURE 1.6. A little rectangular box.

while the surface integral

$$\iint_{\partial V} \langle \mathbf{a} \cdot \mathbf{n} \rangle \, dS \approx$$
$$\approx (a_1^+ - a_1^-)\,\Delta y\,\Delta z + (a_2^+ - a_2^-)\,\Delta x\,\Delta z + (a_3^+ - a_3^-)\,\Delta x\,\Delta y.$$

Here, a_1^+ is the value of component a_1 on the one face in the x direction, and a_1^- its value on the other face, and so on. For example, a_3^+ is the value of a_3 on the box's upper and a_3^- on its lower face. A box has of course six faces, in each of three co-ordinate directions both "up- and downstream".

Then

$$a_1^+ - a_1^- \approx \frac{\partial a_1}{\partial x}\Delta x, \quad a_2^+ - a_2^- \approx \frac{\partial a_2}{\partial y}\Delta y, \quad a_3^+ - a_3^- \approx \frac{\partial a_3}{\partial z}\Delta z,$$

and by substitution we see that

$$\iint_{\partial V} \langle \mathbf{a} \cdot \mathbf{n} \rangle \, dS \approx$$
$$\approx \frac{\partial a_1}{\partial x}\Delta x \cdot \Delta y\,\Delta z + \frac{\partial a_2}{\partial y}\Delta y \cdot \Delta x\,\Delta z + \frac{\partial a_3}{\partial z}\Delta z \cdot \Delta x\,\Delta y =$$

$$= \left(\frac{\partial a_1}{\partial x} + \frac{\partial a_2}{\partial y} + \frac{\partial a_3}{\partial z} \right) \Delta x \, \Delta y \, \Delta z,$$

the same expression as 1.21. So, in this simple case, the Gauss divergence theorem applies.

Obviously the equation also works, if we build out of these "Lego™ bricks" a larger body, because the faces of the bricks touching each other are oppositely oriented and cancel from the surface integral of the whole body. It is a little harder to prove that the equation also applies to bodies having inclined surfaces.

1.12.5 *Example 2: The Poisson equation for a sphere*

The Poisson equation 1.14 states

$$\Delta V = -4\pi G \rho. \tag{1.14}$$

Assume a sphere of radius R, within which the mass density ρ is constant. The volume integral over the sphere gives

$$\iiint_{\mathcal{V}} \Delta V \, d\mathcal{V} = -4\pi G \rho \iiint_{\mathcal{V}} d\mathcal{V} = -4\pi G \rho \mathcal{V} = -4\pi G M, \tag{1.22}$$

in which $M = \rho \mathcal{V}$ is the total mass of the sphere.

On the surface of the sphere, the normal derivative is

$$\frac{\partial V}{\partial n} = \frac{\partial}{\partial r} \frac{GM}{r} \Big|_{r=R} = -\frac{GM}{R^2},$$

a constant, and its integral over the surface of the sphere is

$$\iint_{\partial \mathcal{V}} \frac{\partial V}{\partial n} \, dS = -\frac{GM}{R^2} \cdot S = -\frac{GM}{R^2} \cdot 4\pi R^2 = -4\pi G M. \tag{1.23}$$

Results 1.23 and 1.22 are identical, as Gauss divergence theorem 1.20 requires.

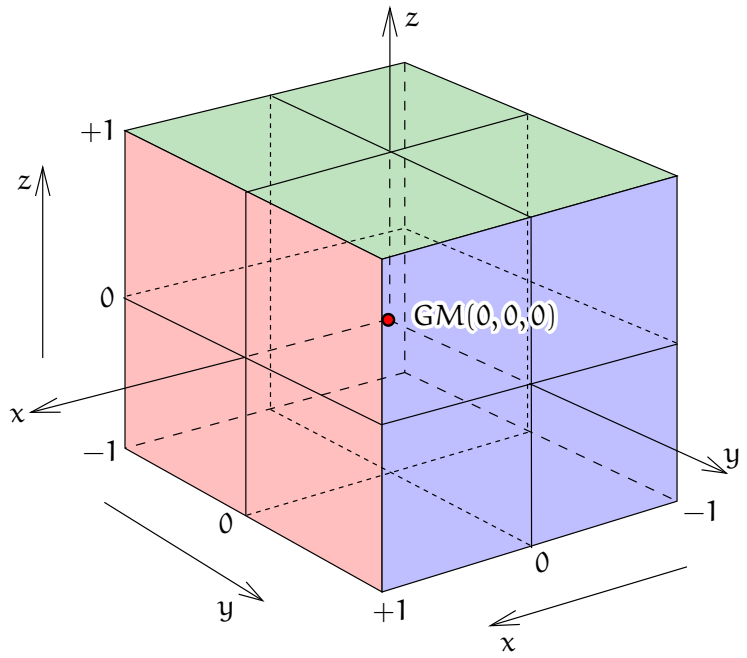


FIGURE 1.7. Eight-unit cube.

1.12.6 *Example 3: A point mass in an eight-unit cube*

See figure 1.7. Assume that there is a point mass of size GM in the centre of a cube bounded by the co-ordinate planes $x = \pm 1$, $y = \pm 1$, and $z = \pm 1$. In that case, the volume integral is

$$\iiint_{\mathcal{V}} \Delta V \, d\mathcal{V} = -4\pi GM \iiint_{\mathcal{V}} \delta(\mathbf{r}) \, d\mathcal{V} = -4\pi GM,$$

in which $\delta(\mathbf{r})$ is Dirac's[18] delta function in space, having an infinite spike at the origin, being zero elsewhere, and producing a value of 1 upon volume integration.

[18]Paul Adrien Maurice Dirac (1902–1984) was an English quantum physicist who found the relativistic wave equation for the electron and theoretically predicted the existence of antimatter. He was a physics Nobel laureate 1933, shared with Erwin Schrödinger. He is also believed to have been on the autism spectrum (Farmelo, 2011).

The surface integral is six times that over the top face

$$\iint_{\text{top face}} \langle \mathbf{a} \cdot \mathbf{n} \rangle \, dS = -GM \int_{-1}^{+1} \left(\int_{-1}^{+1} \frac{1}{(x^2+y^2+1)^{3/2}} \, dx \right) dy.$$

Integrating with respect to x (the expression in large parentheses) yields

$$\int_{-1}^{+1} \frac{1}{(x^2+y^2+1)^{3/2}} \, dx = \left[\frac{x}{(y^2+1)\sqrt{x^2+y^2+1}} \right]_{-1}^{+1} = \\ = \frac{2}{(y^2+1)\sqrt{y^2+2}}.$$

Integrating this with respect to y yields

$$\int_{-1}^{+1} \frac{2}{(y^2+1)\sqrt{y^2+2}} \, dy = \left[2 \arctan \frac{y}{\sqrt{y^2+2}} \right]_{-1}^{+1} = \\ = 4 \arctan \frac{1}{\sqrt{3}} = 4 \cdot \frac{\pi}{6} = \tfrac{2}{3}\pi.$$

Adding the six faces together yields

$$-6 \cdot GM \int_{-1}^{+1} \left(\int_{-1}^{+1} \frac{1}{(x^2+y^2+1)^{3/2}} \, dx \right) dy = -6 \cdot GM \cdot \tfrac{2}{3}\pi = \\ = -4\pi GM,$$

agreeing with the volume result above.

1.13 Green's theorems

Apply the Gauss divergence theorem to the vector field

$$\mathbf{F} = U \, \nabla V.$$

Here, U and V are two different scalar fields. We obtain

$$\iiint_v \operatorname{div} \mathbf{F} \, dv = \iiint_v \langle \nabla \cdot (U \, \nabla V) \rangle \, dv = \\ = \iiint_v U \, \langle \nabla \cdot \nabla \rangle \, V \, dv + \iiint_v \langle \nabla U \cdot \nabla V \rangle \, dv = \\ = \iiint_v U \, \Delta V \, dv + \iiint_v \left(\frac{\partial U}{\partial x}\frac{\partial V}{\partial x} + \frac{\partial U}{\partial y}\frac{\partial V}{\partial y} + \frac{\partial U}{\partial z}\frac{\partial V}{\partial z} \right) dv$$

and

$$\iint_{\partial\mathcal{V}} \langle \mathbf{F}\cdot\mathbf{n}\rangle \,\mathrm{d}S = \iint_{\partial\mathcal{V}} \langle U\,\nabla V\cdot\mathbf{n}\rangle \,\mathrm{d}S = \iint_{\partial\mathcal{V}} U\,\langle \nabla V\cdot\mathbf{n}\rangle \,\mathrm{d}S = \\ = \iint_{\partial\mathcal{V}} U\frac{\partial V}{\partial n}\mathrm{d}S.$$

The result is *Green's*[19] *first theorem*:

$$\iiint_{\mathcal{V}} U\,\Delta V\,\mathrm{d}\mathcal{V} + \iiint_{\mathcal{V}} \left(\frac{\partial U}{\partial x}\frac{\partial V}{\partial x} + \frac{\partial U}{\partial y}\frac{\partial V}{\partial y} + \frac{\partial U}{\partial z}\frac{\partial V}{\partial z}\right)\mathrm{d}\mathcal{V} = \\ = \iint_{\partial\mathcal{V}} U\frac{\partial V}{\partial n}\mathrm{d}S.$$

This may be cleaned up, because the second term on the left is *symmetric* for the interchange of the scalar fields U and V. Let us therefore interchange U and V, and subtract the equations obtained from each other. The result is Green's *second theorem*:

$$\iiint_{\mathcal{V}} (U\,\Delta V - V\,\Delta U)\,\mathrm{d}\mathcal{V} = \iint_{\partial\mathcal{V}} \left(U\frac{\partial V}{\partial n} - V\frac{\partial U}{\partial n}\right)\mathrm{d}S.$$

We assume in all operations that the functions U and V are "well-behaved": for example, all required derivatives exist everywhere in body $\mathcal{V}$.

A useful special case arises by choosing for the function U:

$$U = \frac{1}{\ell},$$

in which ℓ is the distance from the given point of evaluation P. This function U is well-behaved everywhere *except* precisely at point P, where it is not defined.

In the case when point P is outside the surface $\partial\mathcal{V}$, the result, *Green's third theorem*, is now obtained by substitution (remember that $\Delta U = 0$ inside $\partial\mathcal{V}$):

$$\iiint_{\mathcal{V}} \frac{1}{\ell}\Delta V\,\mathrm{d}\mathcal{V} = \iint_{\partial\mathcal{V}} \left(\frac{1}{\ell}\frac{\partial V}{\partial n} - V\frac{\partial}{\partial n}\left(\frac{1}{\ell}\right)\right)\mathrm{d}S.$$

[19] George Green (1793–1841) was a British mathematical physicist, an autodidact, working as a miller near Nottingham. He also invented the word "potential". Green (1828); O'Connor and Robertson (1998); Green's Windmill.

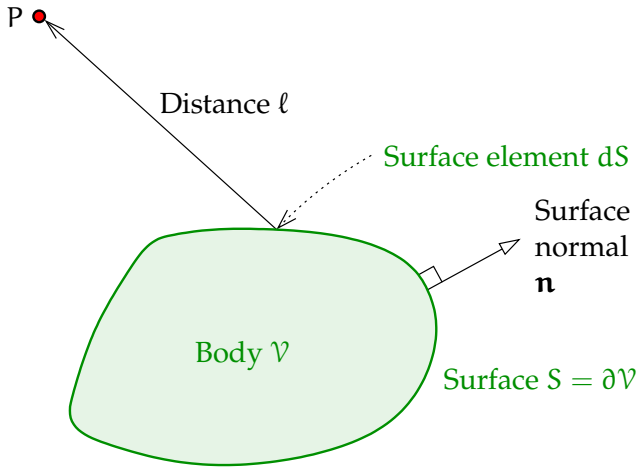


FIGURE 1.8. Geometry for deriving Green's third theorem if point P is outside surface $\partial\mathcal{V}$.

This case is depicted in figure 1.8.

In the case that point P is inside surface $\partial\mathcal{V}$, the computation becomes a little more complicated. Let us learn about a clever technique that — in this case as in others — comes to the rescue.

We form a small sphere of radius ϵ called $\mathcal{V}_2$ around point P; now we can formally define the body (containing a hole) $\mathcal{V} \stackrel{\text{def}}{=} \mathcal{V}_1 - \mathcal{V}_2$, and also its surface $\partial\mathcal{V}$ which consists of two parts, $\partial\mathcal{V} = \partial\mathcal{V}_1 + \partial\mathcal{V}_2$.

Now we may write the volume integral into two parts:

$$\iiint_{\mathcal{V}} \frac{1}{\ell}\Delta V \, d\mathcal{V} = \iiint_{\mathcal{V}_1} \frac{1}{\ell}\Delta V \, d\mathcal{V} - \iiint_{\mathcal{V}_2} \frac{1}{\ell}\Delta V \, d\mathcal{V},$$

where the second term can be integrated in spherical co-ordinates:

$$\iiint_{\mathcal{V}_2} \frac{1}{\ell}\Delta V \, d\mathcal{V} \approx \Delta V_P \int_0^\epsilon 4\pi\ell^2 \frac{1}{\ell} d\ell = 2\pi\Delta V_P \epsilon^2,$$

which will go to zero in the limit $\epsilon \to 0$.

For the first surface integral we obtain using Gauss divergence theorem 1.20 :

$$\iint_{\partial\mathcal{V}_2} \frac{1}{\ell}\frac{\partial V}{\partial n} dS = \frac{1}{\epsilon}\iint_{\partial\mathcal{V}_2} \frac{\partial V}{\partial n} dS = \frac{1}{\epsilon}\iiint_{\mathcal{V}_2} \Delta V \, d\mathcal{V} \approx \frac{1}{\epsilon}\Delta V_P \cdot \tfrac{4}{3}\pi\epsilon^3,$$

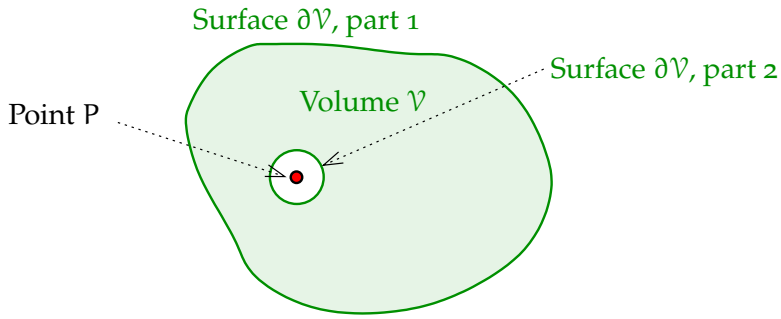


FIGURE 1.9. Geometry for deriving Green's third theorem if point P is inside surface $\partial\mathcal{V}$.

which also goes to zero for $\epsilon \to 0$.

The second surface integral (note that the normal to $\partial\mathcal{V}_2$ is pointing inward to P):

$$\iint_{\partial\mathcal{V}_2} V \frac{\partial}{\partial n}\left(\frac{1}{\ell}\right) dS = \iint_{\partial\mathcal{V}_2} V \cdot -\left(-\frac{1}{\epsilon^2}\right) dS \approx 4\pi\epsilon^2 \cdot \frac{1}{\epsilon^2} V_P = 4\pi V_P.$$

By combining all results with their correct algebraic signs we obtain for the case where P is inside surface $\partial\mathcal{V}_1 \sim \partial\mathcal{V}$:

$$\iiint_{\mathcal{V}} \frac{1}{\ell}\Delta V \, d\mathcal{V} = -4\pi V_P + \iint_{\partial\mathcal{V}} \left(\frac{1}{\ell}\frac{\partial V}{\partial n} - V\frac{\partial}{\partial n}\left(\frac{1}{\ell}\right)\right) dS. \qquad (1.24)$$

After this it must be intuitively clear, and we present without formal proof, that

$$\iiint_{\mathcal{V}} \frac{1}{\ell}\Delta V \, d\mathcal{V} = -2\pi V_P + \iint_{\partial\mathcal{V}} \left(\frac{1}{\ell}\frac{\partial V}{\partial n} - V\frac{\partial}{\partial n}\left(\frac{1}{\ell}\right)\right) dS$$

if point P is precisely *on* the boundary surface $\partial\mathcal{V}$ of body $\mathcal{V}$. This however presupposes that the normal derivative, and especially the *normal direction*, actually exist at precisely point P!

In geodesy, the typical situation is that in which the body $\mathcal{V}$ over the volume of which one wants to evaluate the volume integral, is *the whole space outside the Earth*. In this case, we conveniently have $\Delta V = 0$ and the whole volume integral appearing above vanishes.

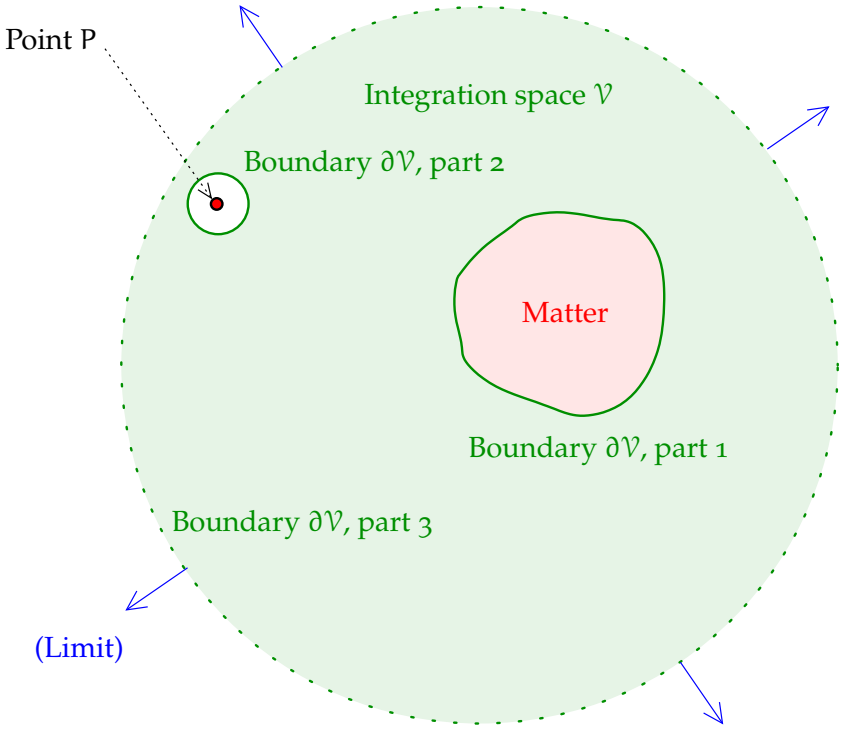


FIGURE 1.10. Green's third theorem for the space external to a body.

Result 1.24 may be generalised to this case, where $\mathcal{V}$ is the whole space outside surface $\partial\mathcal{V}$. This generalisation is made by now choosing as the surface $\partial\mathcal{V}$ the three-part surface $\partial\mathcal{V} = \partial\mathcal{V}_1 + \partial\mathcal{V}_2 + \partial\mathcal{V}_3$, in which $\partial\mathcal{V}_3$ is a sphere of large radius containing both the material body and point P. Its radius is then allowed to *grow* in the limit *to infinity*, and it can be shown that both integrals over the surface $\partial\mathcal{V}_3$ vanish.

The end result is — note that n is now the exterior normal to the Earth's surface:

$$\iiint_{\mathcal{V}} \frac{1}{\ell}\Delta V \, d\mathcal{V} = -4\pi V_P - \iint_{\partial\mathcal{V}} \left(\frac{1}{\ell}\frac{\partial V}{\partial n} - V\frac{\partial}{\partial n}\left(\frac{1}{\ell}\right)\right) dS, \qquad (1.25)$$

Because in this limit, in which $\mathcal{V}$ is the entire empty space exterior to the Earth where $\Delta V = 0$, the left-hand side volume integral vanishes, we may express the potential V_P at point P suitably as a two-term surface integral over surface $\partial\mathcal{V}$.

1.14 The Chasles theorem

We study the above-described case where the "body" is the space outside the surface $\partial\mathcal{V}$ — in practice: the space outside the Earth.

From Green's theorem 1.25 derived above, we may derive for a *harmonic function* V (so, $\Delta V = 0$) in the exterior space:

$$V_P = -\frac{1}{4\pi}\iint_{\partial\mathcal{V}} \frac{1}{\ell}\frac{\partial}{\partial n}V\,dS + \frac{1}{4\pi}\iint_{\partial\mathcal{V}} V\frac{\partial}{\partial n}\left(\frac{1}{\ell}\right)dS. \tag{1.26}$$

Interpretation The exterior, harmonic potential of an arbitrarily shaped body can be represented as the sum of a single and a double mass-density layer on the body's surface.

Explanation We obtain the surface density of a single mass layer using equation 1.15,

$$\kappa = -\frac{1}{4\pi G}\frac{\partial}{\partial n}V, \tag{1.27}$$

and similarly the surface density of a double mass layer using equation 1.18,

$$\mu = \frac{V}{4\pi G}.$$

If we plug these into equation 1.26, we obtain

$$V_P = G\iint_{\partial\mathcal{V}}\left(\frac{\kappa}{\ell} + \mu\frac{\partial}{\partial n}\left(\frac{1}{\ell}\right)\right)dS.$$

In the case that the surface $\partial\mathcal{V}$ is an equipotential surface of the potential V, so $V = V_0$, it follows that a single mass-density layer suffices, because in that case

$$\iint_{\partial\mathcal{V}} V\frac{\partial}{\partial n}\left(\frac{1}{\ell}\right)dS = V_0\iint_{\partial\mathcal{V}}\frac{\partial}{\partial n}\left(\frac{1}{\ell}\right)dS = 0.$$

The right-hand side integral vanishes based on the Gauss divergence theorem. This is because the function $1/\ell$, with ℓ the distance from point P, is *harmonic inside* the Earth's body. The surface of the Earth is $\partial\mathcal{V}$.

20 This is the Chasles[20] theorem, also called the Green equivalent-layer theorem.

21 The theorem is used in Molodensky's[21] theory. The representation of the Earth's gravity field by underground point-mass layers, for example Vermeer (1984), could also be justified with this theorem.

The case where $\partial\mathcal{V}$ is an equipotential surface is realised if the body is fluid and seeks by itself an external form equal to an equipotential surface. For planet Earth, this applies for the ocean surface. In electrostatic theory, for a conductor in which the electrons can move freely, the physical surface will also become an equipotential surface. And the
22 electric charges of a conductor are always on its outer surface.[22]

Equation 1.26, with substitution 1.27, simplifies in this case to

$$V_P = -\frac{1}{4\pi}\iint_{\partial\mathcal{V}} \frac{1}{\ell}\frac{\partial}{\partial n}V \,dS = G\iint_{\partial\mathcal{V}} \frac{\kappa}{\ell}\,dS. \tag{1.28}$$

The equation tells us that we can compute the whole potential exterior to the Earth, if only on the surface of the Earth — the shape of which must be known in order to compute $1/\ell$ — is given the gradient of the potential in the normal or vertical direction $\frac{\partial}{\partial n}V$. This gradient is precisely the gravitational acceleration, a quantity obtainable from gravimetric measurements. All of gravimetric geopotential determination ("geoid determination"), ever since G. G. Stokes, has been based on this idea.

1.15 Boundary-value problems

reuna-arvo-tehtävä The boundary-value problem (BVP) is the problem of computing the

[20] Michel Chasles (1793–1880) was a French mathematician and geometrician, one of the 72 whose names are inscribed on the Eiffel Tower, Eiffel Tower, 72 names.

[21] Mikhail Sergeevich Molodensky (1909–1991) was an illustrious Russian physical geodesist.

[22] This is because the electrostatic potential inside a conductor must also be constant. Even a single extra electron inside the body would make this impossible.

potential V throughout the part of space exterior or interior to a given boundary surface from given values relating to V on that boundary surface, for example on the surface of the Earth. The simplest boundary-value problem is Dirichlet's[23] problem: the potential V itself is given on the boundary surface. More complicated boundary-value problems are based on *linear functionals* of the potential: on the boundary, some linear expression in V is given, for example a derivative or linear combination of derivatives, generally

$$L\{V\},$$

with $L\{\cdot\}$ being a linear functional, see section 10.2.

The Dirichlet boundary-value problem *in the form used in geodesy* is: determine the potential field V if its values are given on a closed surface S, and V is *harmonic* ($\Delta V = 0$) outside surface S. In the vacuum of space, the potential is always harmonic, as already noted earlier: the potential of a point mass m_P, $V = Gm_P/\ell$, is a harmonic function everywhere except at point P — and an extended body consists, in the limit, of many point masses or mass elements.

In the general case, this is a theoretically challenging problem. The existence and uniqueness of the solution has been proven very generally, see Heiskanen and Moritz (1967) page 18.

Based on the values of the potential function V on the surface S we may thus compute the harmonic function $V(x, y, z)$ throughout space outside the surface. The boundary-value problem is a powerful general method also applied in physical geodesy. One must however note that from potential values given on the surface it is not possible to uniquely resolve *the mass distribution inside the Earth*, which generates this potential.

This is clear already in the simple case of a constant potential on the surface of a sphere. If additionally it is given that the mass distribution

[23] Peter Gustav Lejeune Dirichlet (1805–1859) was a German mathematician also known for his contributions to number theory.

is spherically symmetric, then nevertheless the density profile along the radius remains indeterminate. All mass may be concentrated in the centre, or it may form a thin layer just below the sphere's surface, or any alternative between these two extremes. Without additional information — for example from seismic studies or geophysical density models — we cannot resolve this issue.

The Chasles theorem mentioned above, equation 1.26, and its special case, equation 1.28, are also examples of this: the theorem shows how one may describe the exterior potential as generated by a mass distribution on the surface of a body, although we *know* that the field's source is a mass distribution extending throughout the body!

This is a fundamental limitation of all methods that try to obtain information on the situation inside the Earth based only on *gravimetric* measurements on or outside the Earth.

Self-test questions

1. Which instrument was used to determine the constant G? Why is it difficult to obtain a precise value for this constant?
2. Why do all objects, irrespective of their mass, undergo the same acceleration of free fall, although the gravitational attraction on a more massive body is obviously stronger?
3. What is a conservative force field?
   (a) A force field for which the force can be written as the gradient of a unique potential.
   (b) A force field in which an object carried along a closed loop will not gain and not lose energy.
   (c) An attractive force field from which no object can escape.
   (d) A force field the curl of which vanishes everywhere.
4. On the surface of a homogeneous, spherical asteroid the accel-

eration of free fall is $1\,\mathrm{cm/s^2}$. What is the acceleration of free fall on another asteroid that is otherwise similar, but has twice the diameter?

(a) $0.25\,\mathrm{cm/s^2}$

(b) $1\,\mathrm{cm/s^2}$

(c) $2\,\mathrm{cm/s^2}$

(d) $4\,\mathrm{cm/s^2}$

5. What is a harmonic potential?
6. What is the *order* of the Laplace differential equation? kertaluku
7. Is a linear potential, $V(x, y, z) = a + bx + cy + dz$ (a, b, c, d constants), harmonic?
8. If the potential in the previous question is a gravitational potential, calculate its acceleration vector.
9. Under what condition is it possible to describe the external gravitational field emanating from a body as produced by a single mass-density layer on the surface of that body?
10. The dipole-layer density μ is mentioned in section 1.11. What is the SI unit of this quantity?

Exercise 1–1: Core of the Earth

1. Derive the equation giving the acceleration of gravity g on the surface of a homogeneous-density sphere, if the density ρ and radius R_{core} are given.
2. The Earth's iron-nickel core has a mean density of $11\,\mathrm{g/cm^3}$ and its radius is 3500 km. Compute the acceleration of gravitation on its surface g_{core}.
3. What is the attraction g at the centre of the core? What can you say *in general* about the geopotential in this point (*do not try to calculate it*)?

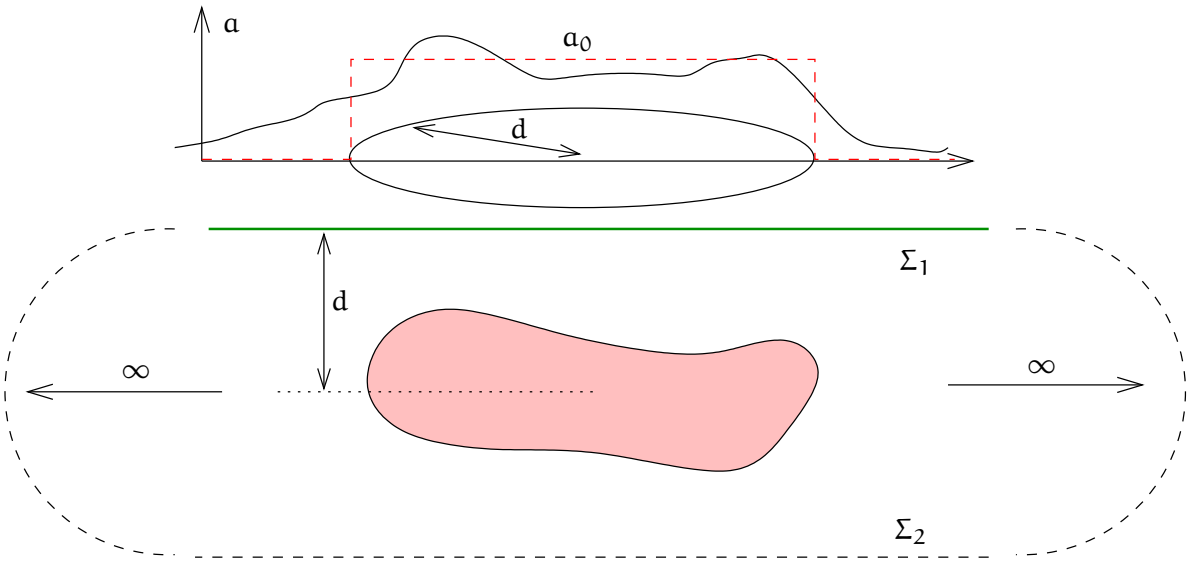


FIGURE 1.11. Iron ore body.

4. Derive the equation for the radial gravitational gradient $\frac{\partial}{\partial r} g$ on the surface of a homogeneous-density sphere of density ρ.

Exercise 1–2: Atmosphere

1. The mean pressure of the atmosphere at sea level is 1013.25 hPa (the unit for pressure, the pascal, is defined as $\mathrm{Pa} = {}^{\mathrm{N}}/_{\mathrm{m}^2}$.) On the Earth's surface gravity is $9.81\,{}^{\mathrm{m}}/_{\mathrm{s}^2}$. Calculate the mean surface density as a thin layer κ in units of ${}^{\mathrm{kg}}/_{\mathrm{m}^2}$.
2. Calculate the total mass of the atmosphere using the spherical-shell approximation. You may take as its radius 6371 km.
3. Calculate the attraction generated by the atmosphere outside it, both as acceleration and as a fraction of the total Earth attraction.
4. How much is the attraction from the atmosphere inside the atmosphere?

Exercise 1–3: The Gauss divergence theorem

There is a deposit (body) of iron ore inside the Earth, figure 1.11. The deposit generates an attraction on the Earth's surface, which has been drawn as the a curve. We use the flat Earth approximation.

The true attraction curve is *approximated* by a simple function

$$a = \begin{cases} a_0 & \text{if } s \leqslant d \\ 0 & \text{if } s > d \end{cases}$$

(red dashed line), where s is the distance from the point on the Earth's surface straight above the ore deposit. So, the area where $a \neq 0$ is a *disk* of radius d on the surface of the Earth.

1. Compute, using the above approximation for a, the surface integral
$$\iint_{\Sigma_1} a \, dS,$$
where Σ_1 is the surface of the Earth, see figure 1.11.

2. According to the Gauss divergence theorem
$$\iint_{\Sigma_1} \langle \mathbf{a}_1 \cdot \mathbf{n}_1 \rangle \, dS + \iint_{\Sigma_2} \langle \mathbf{a}_2 \cdot \mathbf{n}_2 \rangle \, dS = \iiint_{\text{volume}} \Delta V \, d\mathcal{V} = \\ = \iiint_{\text{volume}} (-4\pi G \rho_{\text{iron}}) \, d\mathcal{V} = -4\pi G M_{\text{body}},$$
in which $\Sigma_1 + \Sigma_2$ is the (two part) closed surface around the body. The parts meet at infinity. $\mathbf{a}_1$ and $\mathbf{a}_2$ are the attraction vectors on the surface of the Earth and on the surface Σ_2, and $\mathbf{n}_1$ and $\mathbf{n}_2$ are the exterior normals to the surfaces.
Assuming that
$$\iint_{\Sigma_1} \langle \mathbf{a}_1 \cdot \mathbf{n}_1 \rangle \, dS = \iint_{\Sigma_2} \langle \mathbf{a}_2 \cdot \mathbf{n}_2 \rangle \, dS = -\iint_{\Sigma_1} a \, dS,$$
calculate GM_{body}. Be careful with the algebraic signs!

3. Assuming that the deposit is a sphere at depth d, calculate GM using Newton's law of gravitation from the value a_0 straight above the deposit at the Earth's surface.

4. Compare results 2 and 3 and draw conclusions. Is the function a given above a good approximation?

The Laplace equation and its solutions

2

2.1 The nature of the Laplace equation

The Laplace equation is central to the study of the Earth's gravitational field:

$$\Delta V = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) V = 0.$$

We call the symbol Δ the *Laplace operator*. Sometimes the alternative notation ∇^2 is used.

If we study gravitation as a field, the Laplace equation is more natural than Newton's approach. Newton's equation is used when the mass distribution is known: it yields directly the gravitational force caused by the masses.

The Laplace equation, on the other hand, is a partial differential equation. Its solution gives the *potential* $V(x, y, z)$ of the gravitational field throughout space or a part of space. From this potential one may then calculate the effect of the field on a body moving in space at the location where the body is. This is a two-phase process. It is conceptually new here that a certain property, a *field*, is attached to empty space, and we no longer talk about action at a distance directly between two bodies.

Solving the Laplace equation in the general case may be difficult. The approach generally taken is that we choose some co-ordinate frame

— a rectangular frame (as above), spherical co-ordinates, cylindrical co-ordinates, toroidal co-ordinates, or whatever — which agrees best with the geometry of the problem at hand. Then, we *transform* the Laplace equation to those co-ordinates, we find special solutions of a certain form, and finally we compose a general — or not-so-general — solution as a linear combination of those special solutions: a *series expansion*.

Fortunately, the theory of linear partial differential equations is well-developed. Similar theoretical problems are encountered in the theories of electromagnetic fields (Maxwell theory) and quantum mechanics (Schrödinger[1] equation), not to mention fluid and heat flow.

It is important to note that the Laplace equation is *linear*. This means that, if two solutions are given

$$\Delta V_1 = 0, \qquad \Delta V_2 = 0,$$

then their linear combinations

$$V = \alpha V_1 + \beta V_2, \qquad \alpha, \beta \in \mathbb{R}$$

are also valid solutions: $\Delta V = 0$. This property of linearity makes it possible to seek general solutions as linear combinations or series expansions of basic solutions.

A peculiarity that also distinguishes the Laplace equation from Newton's equation is that it is a *local* equation. It characterises the behaviour of the potential field in a small neighbourhood of one point. However, the solution is sought for a whole area. The solution approach commonly used is the *boundary-value problem*. This means that the field

reuna-arvo-tehtävä

[1] Erwin Rudolf Josef Alexander Schrödinger (1887–1961) was a German physicist and quantum theorist, the inventor of the wave equation of matter named after him, which earned him the 1933 physics Nobel (shared with Paul Dirac), and of the eponymous unobserved cat, which finds itself in a superposition state of being both alive and dead.

values ("boundary values") have to be given only on the boundary of the part of space of interest.

For example, field values are given on the Earth's surface. From this, one calculates the values of the field in outer space, where the Laplace equation applies — the behaviour of the field inside the Earth remains outside the scope of our interest. From the perspective of the exterior gravitational field, one does not even need to know the precise mass distribution inside the Earth — and one cannot even determine it using only measurement values obtained on and above the Earth's surface.

2.2 The Laplace equation in rectangular co-ordinates

It is a learning experience to write and solve the Laplace equation in rectangular co-ordinates. The case is analogous to that of spherical co-ordinates but the maths is much simpler.

Assume that the Earth's surface, sea level, is the level surface for z co-ordinates $z = 0$. Write

$$\Delta V = \Delta\big(V(x,y,z)\big) = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)\big(X(x)\cdot Y(y)\cdot Z(z)\big),$$

containing a *trial solution*

$$V(x,y,z) = X(x)\cdot Y(y)\cdot Z(z).$$

In other words, we write experimentally V as the product of three factor functions, with each factor function depending only on one co-ordinate — the "separation of variables". A realistic potential function V will of course usually not be in this form. We may however hope to write it as a linear combination of terms that *are* in the above form, thanks to the linearity of the Laplace equation.

muuttujien erottaminen

By taking all the partial derivatives, we obtain

$$YZ\frac{\partial^2}{\partial x^2}X + XZ\frac{\partial^2}{\partial y^2}Y + XY\frac{\partial^2}{\partial z^2}Z = 0.$$

Dividing by the expression XYZ yields

$$\frac{1}{X(x)}\frac{\partial^2}{\partial x^2}X(x) + \frac{1}{Y(y)}\frac{\partial^2}{\partial y^2}Y(y) + \frac{1}{Z(z)}\frac{\partial^2}{\partial z^2}Z(z) = 0.$$

Because this has to be true throughout the space, that is for *all* combinations of values x, y, and z, it follows that *each term must be a constant*. If we take for the first and second constants $-\omega_x^2$ and $-\omega_y^2$, we get in conclusion for the third constant $\omega_x^2 + \omega_y^2$. By writing this definition and result out and moving the denominator to the other side, we obtain

$$\frac{\partial^2}{\partial x^2}X(x) = -\omega_x^2 X(x), \quad \frac{\partial^2}{\partial y^2}Y(y) = -\omega_y^2 Y(y),$$

(the reason for choosing a negative constant will become apparent), and

$$\frac{\partial^2}{\partial z^2}Z(z) = \left(\omega_x^2 + \omega_y^2\right) Z(z).$$

Now, the solution is readily found at least to the first two equations: they are *harmonic oscillators*, and their basis solutions[2] are

[2] harmoninen värähtelijä

$$X(x) = \exp\left(\pm i\omega_x x\right), \quad Y(y) = \exp\left(\pm i\omega_y y\right).$$

The solution of the Z equation, on the other hand, is exponential:

$$Z(z) = \exp\left(\pm z\sqrt{\omega_x^2 + \omega_y^2}\right).$$

We can now form basis solutions in space:

$$V_{\omega_x\omega_y}(x, y, z) = \exp\left(i\left(\pm\omega_x x \pm \omega_y y\right) \pm z\sqrt{\omega_x^2 + \omega_y^2}\right).$$

The general solution is obtained by summing the terms $V_{\omega_x\omega_y}$ for different values of ω_x and ω_y with varying coefficients.

We cannot choose the value pair (ω_x, ω_y) entirely freely. The values which are allowed will depend on the *boundary conditions* given.

Let us assume that in both the x and y directions the size of our world is L ("shoebox world"[3]). Let us make things a little simpler by assuming that on the boundary surfaces of our shoebox world we have the *boundary conditions*

$$V(0, y, z) = V(L, y, z) = V(x, 0, z) = V(x, L, z) = 0.$$

It then follows that the only pairs (ω_x, ω_y) that yield a solution that fits the box are

$$\omega_x = \frac{\pi j}{L}, \quad \omega_y = \frac{\pi k}{L}, \qquad j, k \in \mathbb{Z},$$

and the only suitable functions are sine functions. Thus we obtain as a solution

$$V_{jk}(x, y, z) = \sin\left(\pi j \frac{x}{L}\right) \sin\left(\pi k \frac{y}{L}\right) \exp\left(\pm\pi\sqrt{(j^2 + k^2)}\frac{z}{L}\right).$$

This particular solution may now be generalised by multiplying it by suitable coefficients, and summing it over different index values $j = 0, \pm1, \pm2, \ldots$ and $k = 0, \pm1, \pm2, \ldots$.

We may however remark that the terms for which $j = 0$ or $k = 0$ will always vanish, and the terms that contain $j = +n$ and $j = -n$, or $k = +n$ and $k = -n$, $n \in \mathbb{N}$, are (apart from their algebraic signs) identical. Therefore in practice we sum over the values $j = 1, 2, \ldots$ and $k = 1, 2, \ldots$.

Different boundary conditions will give slightly different general solutions. Their general form is, however, always similar.

The zero-level $z = 0$ expansion resulting from the general solution is the familiar Fourier[4] sine expansion:

---

[2] Alternative basis solutions are $X(x) = \sin \omega_x x$, $X(x) = \cos \omega_x x$ etc. They are equivalent to those presented because $\exp(i\omega_x x) = \cos \omega_x x + i \sin \omega_x x$ and $\exp(-i\omega_x x) = \cos \omega_x x - i \sin \omega_x x$.

[3] ... although real-world shoeboxes are rarely square.

$$V(x,y,0) = \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} v_{jk} \overbrace{\sin\left(\pi\frac{jx}{L}\right)\sin\left(\pi\frac{ky}{L}\right)}^{V_{jk}(x,y)}, \tag{2.1}$$

in which v_{jk} are Fourier coefficients, and the expressions

$$V_{jk}(x,y) \stackrel{\text{def}}{=} \sin\left(\pi\frac{jx}{L}\right)\sin\left(\pi\frac{ky}{L}\right)$$

kantafunktio are two-dimensional basis functions on the Earth's surface, level $z = 0$.

We refer to section B.2.2 in appendix B for a description with illustration of how a Fourier analysis and synthesis on a simple function is done, and how the Fourier expansion approximates the original function as terms are added.

A complete three-dimensional expansion again is

$$V(x,y,z) = \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} v_{jk} \overbrace{\sin\left(\pi\frac{jx}{L}\right)\sin\left(\pi\frac{ky}{L}\right)}^{V_{jk}(x,y)} \exp\left(\pm\pi\sqrt{j^2+k^2}\frac{z}{L}\right). \tag{2.2}$$

Inside the z expression there may be a positive as well as a negative algebraic sign! Of course the solution with a positive sign goes to $\to \infty$ when $z \to \infty$, which is not physically realistic in the exterior space.

Note also that $V(x,y,0)$ and v_{jk} represent *the same gravitational field* in two essentially different ways: in the space domain, and in the — spatial — frequency, or wave-number, domain. The information content in the two is the same. They can be transformed into each other by the forward and inverse Fourier transforms $\mathcal{F}$ and $\mathcal{F}^{-1}$.

In fact, the information content in $V(x,y,0)$ is in principle the same as that in $V(x,y,z)$ for *any* level z: knowing the potential on one surface means — with the Laplace equation — knowing the potential throughout space.

kommutoiva kaavio We summarise equations 2.1 and 2.2 in commutative diagram 2.2.

[4] Joseph Fourier (1768–1830) was a French mathematician and physicist — and some would say, climatologist — one of the Eiffel Tower's 72 names, Eiffel Tower, 72 names.

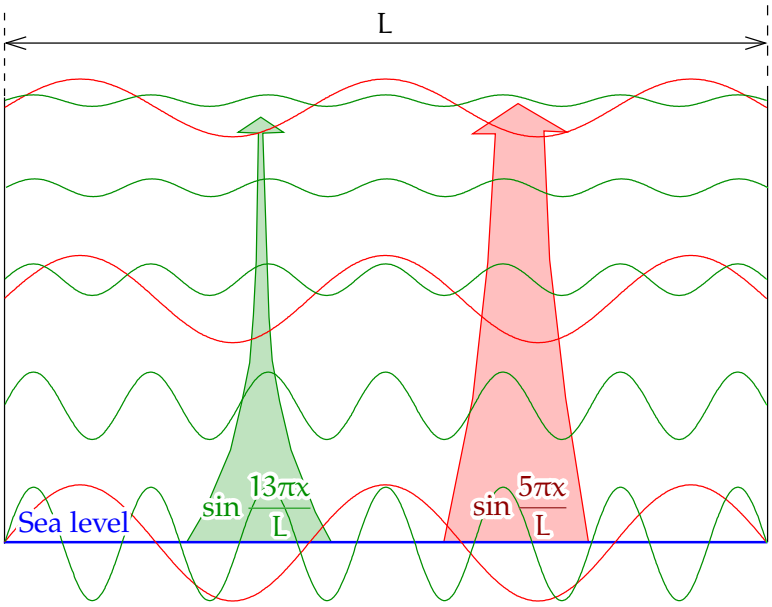


FIGURE 2.1. The exponential attenuation of the Fourier waviness of a harmonic field with height. Rectangular geometry, one horizontal dimension. Long waves (small wave numbers, red) attenuate more slowly than short waves (green), meaning that height acts as a low-pass filter.

The takeaway from this is that the operation of vertically shifting the field V from zero level to the level z, which is not straightforward in the space domain, becomes simple — as in a straightforward multiplication

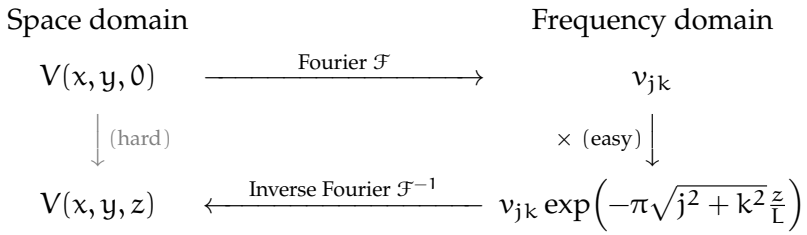


FIGURE 2.2. Vertically shifting the harmonic field V in the space and frequency (wave-number) domains, commutative diagram. Rectangular geometry.

— in the frequency domain.[5] The same applies in spherical co-ordinates, where the frequency domain means spherical-harmonic coefficients, as we shall see.

2.3 The Laplace equation in polar co-ordinates

In polar co-ordinates, two-dimensionally, the Laplace equation is

$$\Delta V = \frac{\partial^2 V}{\partial r^2} + \frac{1}{r}\frac{\partial V}{\partial r} + \frac{1}{r^2}\frac{\partial^2 V}{\partial \alpha^2} = 0.$$

We carry out on this the same kind of separation of variables as was done in section 2.2. Write first

$$V(\alpha, r) = A(\alpha)\, R(r)$$

and then split the above equation into two equations, one for the right-hand side function $R(r)$ and one for the function $A(\alpha)$. Substitution yields

$$A(\alpha)\frac{\partial^2 R(r)}{\partial r^2} + \frac{A(\alpha)}{r}\frac{\partial R(r)}{\partial r} + \frac{R(r)}{r^2}\frac{\partial^2 A(\alpha)}{\partial \alpha^2} = 0.$$

Multiply by the expression $r^2/A(\alpha)\, R(r)$:

$$\left(\frac{r^2}{R(r)}\frac{\partial^2 R(r)}{\partial r^2} + \frac{r}{R(r)}\frac{\partial R(r)}{\partial r}\right) + \frac{1}{A(\alpha)}\frac{\partial^2 A(\alpha)}{\partial \alpha^2} = 0.$$

Both terms must be constant:

$$r\left(r\frac{\partial^2 R(r)}{\partial r^2} + \frac{\partial R(r)}{\partial r}\right) - k^2\, R(r) = 0,$$

$$\frac{\partial^2 A(\alpha)}{\partial \alpha^2} + k^2 A(\alpha) = 0.$$

Here, the algebraic sign of the constant k^2 has been chosen so that $A(\alpha)$ gets a periodic solution. Such a general solution would be

$$A_k(\alpha) = a_k \cos k\alpha + b_k \sin k\alpha,$$

[5] The reason for this, as we shall later discuss more generally, is that the vertical shift operation is a *convolution*.

in which, because angle α has a period of 2π, k has to be a non-negative integer: $k = 0, 1, 2, \ldots$. Negative k does not give different solutions, because

$$a_k \cos k\alpha = a_k \cos(-k\alpha), \qquad b_k \sin k\alpha = -b_k \sin(-k\alpha).$$

The other equation, in the function $R(r)$, is harder to solve. A test solution is a power law:

$$R(r) = r^q.$$

Substitution yields

$$\begin{aligned} r\left(rq\,(q-1)\,r^{q-2} + qr^{q-1}\right) - k^2 r^q = 0 \\ \implies \quad q^2 - k^2 = 0 \\ \implies \quad q^2 = k^2. \end{aligned}$$

This works for positive $q = 2, 3, \ldots$ and negative $q = -1, -2, \ldots$. For $q = 1$ we find

$$r - k^2 r = 0 \quad \implies \quad k^2 = 1 = q^2.$$

For $k = 0$, besides the trivial constant solution, the non-trivial solution $R(r) = \ln r$ is found:

$$r\left(r \cdot -\frac{1}{r^2} + \frac{1}{r}\right) - k^2 \ln r = 0 \quad \implies \quad k = 0.$$

Thus we obtain the general solution

$$R_k(r) = \begin{cases} 1 \text{ or } \ln r & \text{if } k = 0, \\ r^k \text{ or } r^{-k} & \text{if } k = 1, 2, \ldots. \end{cases}$$

We see that, if we require the solution to exist at the origin $r = 0$, we need the *first* solutions, obtaining

$$V^{\text{int}}(\alpha, r) = a_0 + \sum_{k=1}^{\infty} r^k \left(a_k \cos k\alpha + b_k \sin k\alpha\right),$$

but if we require existence — or, at least, good behaviour — at infinity[6] [6]

$r \to \infty$, we need the *second* solutions,

$$V^{\text{ext}}(\alpha, r) = a_0 + b_0 \ln r + \sum_{k=1}^{\infty} r^{-k} \left(a_k \cos k\alpha + b_k \sin k\alpha \right). \tag{2.3}$$

There is a clear similarity here to the three-dimensional, spherical co-ordinates, case.

2.4 Spherical, geodetic, ellipsoidal co-ordinates

In physical geodesy we use geometrical and physical concepts side by side. For example, the co-ordinates of a point can be given in the form (X, Y, Z), which are in principle geometric — except for the physical assumption that the origin of the co-ordinate system is in the centre of mass of the Earth.

pyörähdys-ellipsoidi

As the Earth is not precisely a sphere but rather an oblate ellipsoid of revolution, one cannot use geographical co-ordinates as if they were spherical co-ordinates. Because the flattening of the Earth — some 0.3 % — cannot be ignored, this difference is significant. The connection between spherical co-ordinates (ϕ, λ, r) and rectangular ones (X, Y, Z) is the following:

$$\begin{aligned} X &= r \cos\phi \cos\lambda, \\ Y &= r \cos\phi \sin\lambda, \\ Z &= r \sin\phi. \end{aligned} \tag{2.4}$$

Here ϕ is the geocentric latitude and λ is the (ordinary — geocentric, geodetic or geographical, all three are the same) longitude, while r is the distance from the Earth's centre. The X axis points in the direction of the Greenwich meridian. See figure 2.3.

On the Earth's surface, these spherical co-ordinates are not very useful because of the Earth's flattening, but in space, spherical co-ordinates

[6] In fact, $\lim_{r\to\infty} V^{\text{ext}} \to \infty$ but $\lim_{r\to\infty} \frac{\partial}{\partial r} V^{\text{ext}} = 0$.

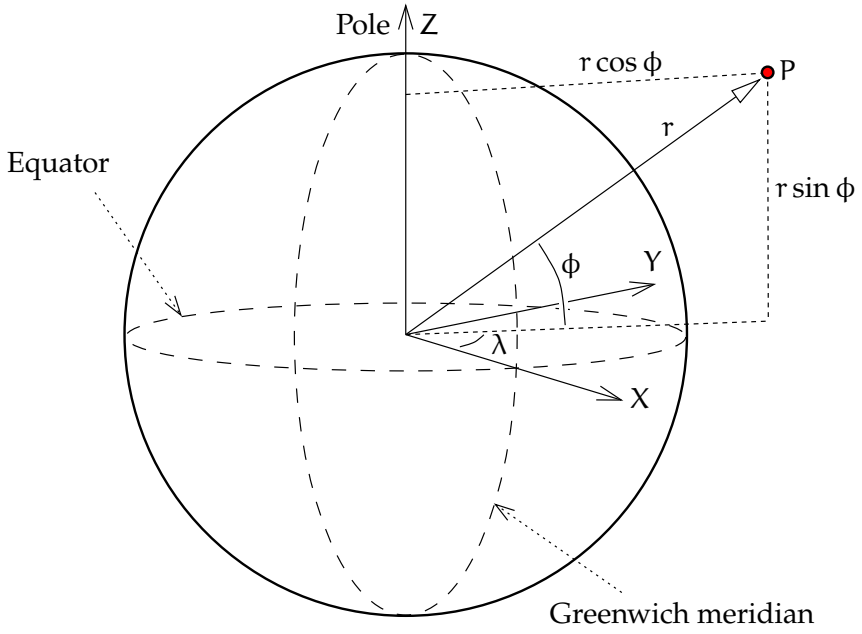


FIGURE 2.3. Definition of spherical co-ordinates.

are much-used. On the Earth's surface most often *geodetic* — also called geographical — co-ordinates (φ, λ, h) are used:

$$\begin{aligned} X &= (N + h)\cos\varphi\cos\lambda, \\ Y &= (N + h)\cos\varphi\sin\lambda, \\ Z &= \left(N + h - e^2 N\right)\sin\varphi, \end{aligned} \tag{2.5}$$

in which

$$N(\varphi) = \frac{a}{\sqrt{1 - e^2\sin^2\varphi}} = \frac{a^2}{\sqrt{a^2\cos^2\varphi + b^2\sin^2\varphi}}. \tag{2.6}$$

The quantity N defined in equation 2.6 is the west-east direction, or *transversal*, radius of curvature of the reference ellipsoid. In the equation, a is the equatorial radius of the reference ellipsoid used, b is the polar radius,

poikittainen vertaus-ellipsoidi

$$e^2 \stackrel{\text{def}}{=} \frac{a^2 - b^2}{a^2} \tag{2.7}$$

is the square of the *first eccentricity*,[7] and in equations 2.5, h is the height

[7]

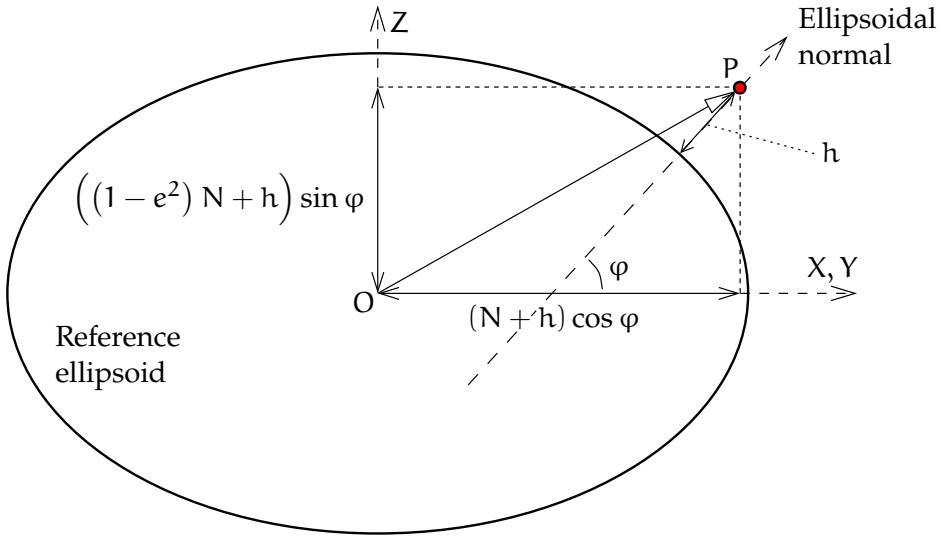


FIGURE 2.4. Definition of geodetic co-ordinates.

of the point above the reference ellipsoid, see figure 2.4.

Converting rectangular co-ordinates into geodetic ones is easiest to do iteratively, although the literature also offers closed formulas.

Spherical co-ordinates and geodetic, also called geographical, co-ordinates are considerably different. In latitude, the difference is up to 11 minutes of arc, or almost 20 kilometres. This maximum is attained for latitudes $\pm 45°$.

In theoretical work one also uses *ellipsoidal co-ordinates* (β, λ, u). The co-ordinate β is called the *reduced latitude*. The relationship with rectangular co-ordinates is

redukoitu leveysaste

$$\begin{aligned} X &= \sqrt{u^2 + E^2}\cos\beta\cos\lambda, \\ Y &= \sqrt{u^2 + E^2}\cos\beta\sin\lambda, \\ Z &= u\sin\beta. \end{aligned} \tag{2.8}$$

isoakselin puolikas

If the semimajor axis of the Earth ellipsoid is a and its semiminor axis

pikkuakselin puolikas

[7] The parameter is connected to the Earth's flattening f through the equation $e^2 = 2f - f^2$.

b, it follows that $E^2 = a^2 - b^2$. Equation 2.7 tells us that $E^2 = a^2e^2$. The first eccentricity e is the eccentricity of the *meridian ellipse*, and $E = ae$ is the distance from the geocentre of the two *focal points* of this ellipse, which lie on both sides of the geocentre in the equatorial plane. Equations 2.8 tell us that all ellipses $u =$ constant in the meridian plane for different values of u have the same two focal points: they are *confocal*. See figure 4.6.

We note still — see Heiskanen and Moritz (1967) figure 1-14 — that the curves $\beta =$ constant describe *hyperbolas* with these same focal points.

2.5 The Laplace equation in spherical co-ordinates

The Laplace equation transformed to spherical co-ordinates reads (see appendix E for a geometric proof):

$$\Delta V = \frac{\partial^2 V}{\partial r^2} + \frac{2}{r}\frac{\partial V}{\partial r} + \frac{1}{r^2}\frac{\partial^2 V}{\partial \phi^2} - \frac{\tan\phi}{r^2}\frac{\partial V}{\partial \phi} + \frac{1}{r^2\cos^2\phi}\frac{\partial^2 V}{\partial \lambda^2} = 0, \quad (2.9)$$

in which ϕ is the (geocentric) latitude, λ is the longitude, and r is the distance from the origin or centre of the Earth.

Here we shall not derive the solution of this equation by separation of variables, as it is pretty complicated. It can be found in section E.2 and in the literature (Heiskanen and Moritz, 1967, section 1–9). What is significant is that the solution looks somewhat similar to the solution in rectangular co-ordinates presented earlier, section 2.2. The basis solutions of the Laplace equation are

$$V_n^{\text{int}}(\phi,\lambda,r) = r^n Y_n(\phi,\lambda), \quad V_n^{\text{ext}}(\phi,\lambda,r) = \frac{Y_n(\phi,\lambda)}{r^{n+1}}, \qquad n = 0, 1, \ldots, \quad (2.10)$$

of which the first is nonphysical in outer space, because, unlike the true geopotential, these expressions grow to infinity for $r \to \infty$.

In the above equations, the functions $Y_n(\phi,\lambda)$ are called *surface spherical harmonics*, whereas the functions $V_n(\phi,\lambda,r)$ are *solid spherical harmonics*. The latter are *harmonic* functions everywhere in space except

pinta-pallofunktio

avaruus-pallofunktio

at the origin (2.10, rightmost equation) or at infinity (leftmost equation).

The functions Y_n, called *Laplace's spherical harmonics*, are

$$Y_n(\phi,\lambda) = \sum_{m=0}^{n} P_{nm}(\sin\phi)\,(a_{nm}\cos m\lambda + b_{nm}\sin m\lambda)\,. \qquad (2.11)$$

The functions P_{nm} are *Legendre functions*, on which more later on. With the help of expression 2.11, we obtain, by using the second, physically realistic alternative from equations 2.10, the following solution or *series expansion* for the potential V in outer space:

pallofunktio-kehitelmä

$$V(\phi,\lambda,r) = \sum_{n=0}^{\infty} \frac{1}{r^{n+1}} \sum_{m=0}^{n} P_{nm}(\sin\phi)\,(a_{nm}\cos m\lambda + b_{nm}\sin m\lambda)\,. \qquad (2.12)$$

The coefficients a_{nm} and b_{nm} are called the coefficients of the spherical-harmonic expansion, in short, *spectral coefficients*. Together they represent the function V, in somewhat the same way that the Fourier coefficients v_{jk} do in rectangular co-ordinates in equation 2.2. The subscripts n and m are called *degree* and *order*.

asteluku, järjestysluku

Often we will be using a somewhat freer notation for the scaled functions Y_n/R^{n+1}. For example, if we expand the disturbing potential T into spherical harmonics, we shall use the notation $T_n(\phi,\lambda)$ for its surface harmonics. Similarly, $\Delta g_n(\phi,\lambda)$ is the surface harmonic of the gravity anomaly Δg for degree n, and so on. Then, it holds on the Earth's surface $r = R$ (*degree constituent expansion*) that

häiriö-potentiaali pallofunktio

asteosuus-hajotelma

$$T(\phi,\lambda,R) = \sum_{n=0}^{\infty} T_n(\phi,\lambda), \qquad \Delta g(\phi,\lambda,R) = \sum_{n=0}^{\infty} \Delta g_n(\phi,\lambda),$$

and so on.

2.6 Dependence on height

From the above equations 2.10 one sees that for different values of the degree n the function $V_n(\phi,\lambda,r)$ has a different dependence on

the distance r from the Earth's centre, or equivalently, on the height $H = r - R$, if by R we denote the radius of the Earth sphere or sea level. The dependence is

$$V_n(\phi, \lambda, r) = \frac{Y_n(\phi, \lambda)}{r^{n+1}}.$$

At sea level

$$V_n(\phi, \lambda, R) = \frac{Y_n(\phi, \lambda)}{R^{n+1}} \stackrel{\text{def}}{=} V_n(\phi, \lambda).$$

Therefore we may write

$$V_n(\phi, \lambda, r) = \left(\frac{R}{r}\right)^{n+1} V_n(\phi, \lambda) = \left(\frac{R+H}{R}\right)^{-(n+1)} V_n(\phi, \lambda) =$$
$$= \left(1 + \frac{H}{R}\right)^{-(n+1)} V_n(\phi, \lambda) \approx \exp\left(-\frac{H}{R}(n+1)\right) V_n(\phi, \lambda).$$

We see that the attenuation of the potential with height is *exponential*, and the harmonic degree number n appears in the exponent, as also did the wave number in rectangular geometry, see equation 2.2 and figure 2.1. The analogy works.

Self-test questions

1. In what essential way does the approach of the Laplace equation differ from Newton's approach?
2. How does the linearity of the Laplace equation help in finding solutions?
3. How does separation of variables work?
4. Why does solving the Laplace equation require boundary conditions?
5. Show, using equations 2.8, that for the curves u = constant in the meridian plane $Y = 0$, the *sum* of the distances from the curve point $\begin{bmatrix} \sqrt{u^2 + E^2}\cos\beta & 0 & \sin\beta \end{bmatrix}^T$ to the focal points $\begin{bmatrix} \pm E & 0 & 0 \end{bmatrix}^T$ is constant (and that these curves thus are confocal ellipses), and that for the curves β = constant the *difference* of these distances is

constant (and that these curves thus are confocal hyperbolas). See figure 4.6.

Legendre functions and spherical harmonics

3

3.1 Legendre functions

In equations 2.11 and 2.12, the functions P are *Legendre*[1] *functions* that pop up whenever we solve a Laplace-like equation in spherical coordinates. There exist various effective, so-called recursive algorithms for their calculation, for example the following algorithm only for ordinary Legendre polynomials $P_n = P_{n0}$: [1]

$$nP_n(t) = -(n-1)\,P_{n-2}(t) + (2n-1)\,tP_{n-1}(t). \qquad (3.1)$$

Similar equations also exist for the functions P_{nm}, $m > 0$. There are even alternatives to choose from, although most equations are complicated. One should be careful that, in their computation, the factorials do not go overboard. Already 30! (factorial of 30) is a larger number than computers can handle as 64-bit integers — not to mention 360!, the factorial of 360, which overflows even the standard 64-bit floating-point format. Heiskanen and Moritz's (1967) equation 1-62, contrary to what is stated there, is *not* directly suitable for computer use! kertoma

[1] Adrien-Marie Legendre (1752–1833) was a French mathematician known for his work on number theory, statistics — he invented the method of least-squares independently from Gauss — and elliptical functions. His name is inscribed on the Eiffel Tower, Eiffel Tower, 72 names.

TABLE 3.1. Legendre polynomials. $t = \sin\phi$.

As a function of t	Expressed in sines and cosines
$P_0(t) = 1$	$P_0(\sin\phi) = 1$
$P_1(t) = t$	$P_1(\sin\phi) = \sin\phi$
$P_2(t) = \frac{3}{2}t^2 - \frac{1}{2}$	$P_2(\sin\phi) = -\frac{3}{4}\cos 2\phi + \frac{1}{4}$
$P_3(t) = \frac{5}{2}t^3 - \frac{3}{2}t$	$P_3(\sin\phi) = -\frac{5}{8}\sin 3\phi + \frac{3}{8}\sin\phi$
$P_4(t) = \frac{1}{8}\left(35t^4 - 30t^2 + 3\right)$	$P_4(\sin\phi) = \frac{35}{64}\cos 4\phi - \frac{5}{16}\cos 2\phi + \frac{9}{64}$
$P_5(t) = \frac{1}{8}\left(63t^5 - 70t^3 + 15t\right)$	
$P_6(t) = \frac{1}{16}\left(231t^6 - 315t^4 + 105t^2 - 5\right)$	

The first Legendre polynomials are listed in table 3.1. Higher polynomials than this are rarely needed in manual computation.

kantafunktio For comparison, the Fourier *basis functions* (like, in a more complicated

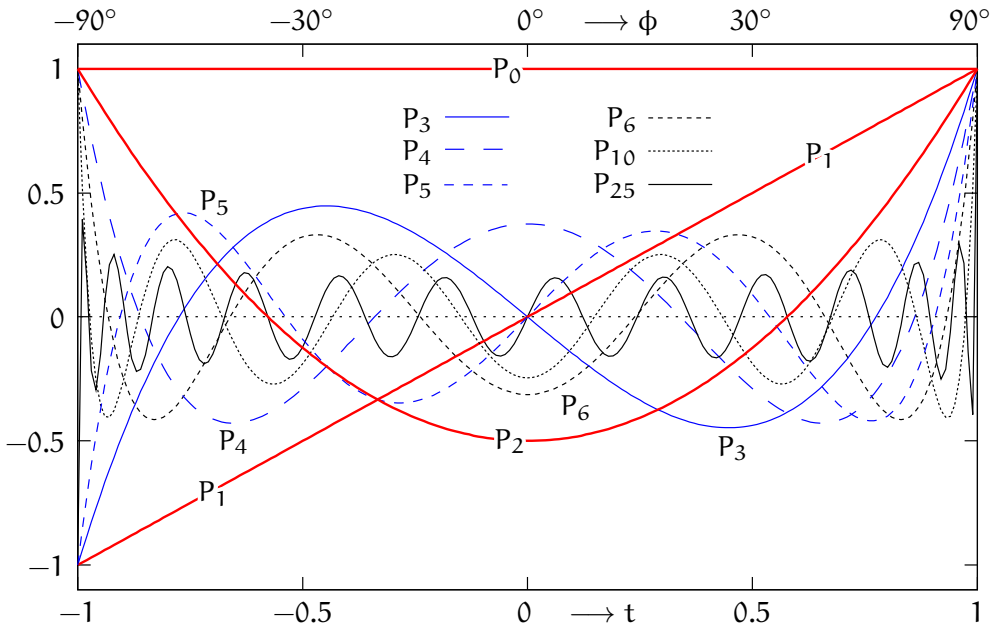


FIGURE 3.1. A number of Legendre polynomials $P_0(t), \ldots, P_{25}(t)$ as functions of the argument $t = \sin\phi$.

way, sines and cosines together as well!)

$$F_j(x) = \exp\left(2\pi i j \frac{x}{L}\right),$$

in which $i^2 = -1$, can also be calculated recursively:

$$F_{j+1}(x) = F_j(x) \cdot F_1(x).$$

3.1.1 *Properties of Legendre polynomials*

- The even polynomials — those polynomials of which the harmonic degree number n is even — are mirror symmetric through the origin $\phi = t = 0$, the equatorial plane: $P_n(-t) = P_n(t)$, or equivalently $P_n\big(\sin(-\phi)\big) = P_n(\sin\phi)$. This means that their values at the same latitude north and south of the equator are identical. The odd polynomials are again antisymmetric: $P_n(-t) = -P_n(t)$ or $P_n\big(\sin(-\phi)\big) = -P_n(\sin\phi)$, meaning that their values at the same latitude north and south of the equator are of opposite signs.
- From figure 3.1, we see that the polynomials $P_n(t)$ go, on the whole interval $t \in [-1, 1]$, or $\phi \in [-90°, 90°]$, precisely n times through zero.
- As the values in the end points $t = \pm 1$, $\phi = \pm 90°$ are ± 1, it follows that there are precisely $n + 1$ "algebraic-sign intervals", meaning open intervals of t or ϕ on which the polynomial assumes only positive or only negative values.

3.1.2 *Properties of associated Legendre functions*

We give several of the associated Legendre functions P_{nm}, $m \neq 0$ in table 3.2 for illustration.

Legendren liitännäisfunktio

One defining equation for these is

$$P_{nm}(t) = \left(1 - t^2\right)^{m/2} \frac{d^m P_n(t)}{dt^m}. \tag{3.2}$$

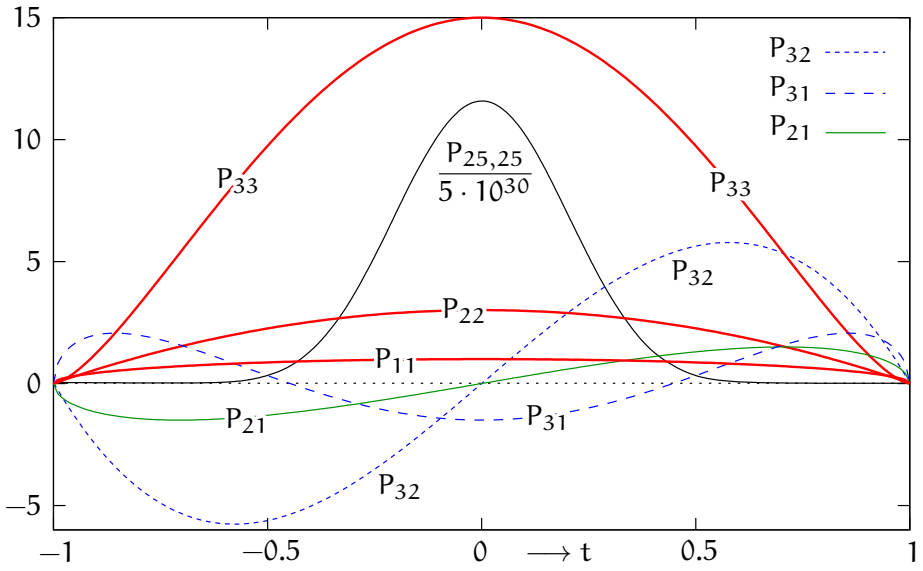


FIGURE 3.2. Some associated Legendre functions. Note the extremely different scale used for the function $P_{25,25}$, see equation 3.8.

- The associated Legendre functions are also either mirror symmetric through the origin or equatorial plane, $P_{nm}(-t) = P_{nm}(t)$, or equivalently, $P_{nm}\big(\sin(-\phi)\big) = P_{nm}(\sin\phi)$, or antisymmetric, $P_{nm}(-t) = -P_{nm}(t)$ or $P_{nm}\big(\sin(-\phi)\big) = -P_{nm}(\sin\phi)$, depending on the values of degree number n and order number m.

asteluku, järjestysluku

- Figure 3.2 suggests that the polynomials $P_{nm}(t)$ go on $t \in [-1, 1]$,

TABLE 3.2. Associated Legendre functions.

Function of t	Trigonometric function
$P_{11}(t) = \sqrt{1-t^2}$	$P_{11}(\sin\phi) = \cos\phi$
$P_{21}(t) = 3t\sqrt{1-t^2}$	$P_{21}(\sin\phi) = 3\sin\phi\cos\phi$
$P_{22}(t) = 3\left(1-t^2\right)$	$P_{22}(\sin\phi) = 3\cos^2\phi$
$P_{31}(t) = \frac{3}{2}\left(5t^2-1\right)\sqrt{1-t^2}$	$P_{31}(\sin\phi) = \frac{3}{2}\left(5\sin^2\phi - 1\right)\cos\phi$
$P_{32}(t) = 15t\left(1-t^2\right)$	$P_{32}(\sin\phi) = 15\sin\phi\cos^2\phi$
$P_{33}(t) = 15\left(1-t^2\right)^{3/2}$	$P_{33}(\sin\phi) = 15\cos^3\phi$

or $\phi \in [-90°, 90°]$, precisely $n - m$ times through zero. This is indeed the case.

- As the values in the end points $t = \pm 1$, $\phi = \pm 90°$ are also zero, it follows that there are precisely $n - m + 1$ "algebraic-sign intervals".

3.1.3 *Surface spherical harmonics*

Starting from equation 2.11, we may write

$$\begin{aligned} Y_n(\phi, \lambda) &= \\ &= \sum_{m=0}^{n} \left(a_{nm} P_{nm}(\sin\phi) \cos m\lambda + b_{nm} P_{nm}(\sin\phi) \sin m\lambda \right) = \\ &= \sum_{m=-n}^{n} \nu_{nm} Y_{nm}(\phi, \lambda), \end{aligned}$$

in which m now runs from $-n$ to $+n$. Here

$$Y_{nm}(\phi, \lambda) \stackrel{\text{def}}{=} \begin{cases} P_{nm}(\sin\phi) \cos m\lambda & \text{if } m \geqslant 0, \\ P_{n|m|}(\sin\phi) \sin |m| \lambda & \text{if } m < 0, \end{cases} \tag{3.3}$$

$$\nu_{nm} \stackrel{\text{def}}{=} \begin{cases} a_{nm} & \text{if } m \geqslant 0, \\ b_{n|m|} & \text{if } m < 0. \end{cases} \tag{3.4}$$

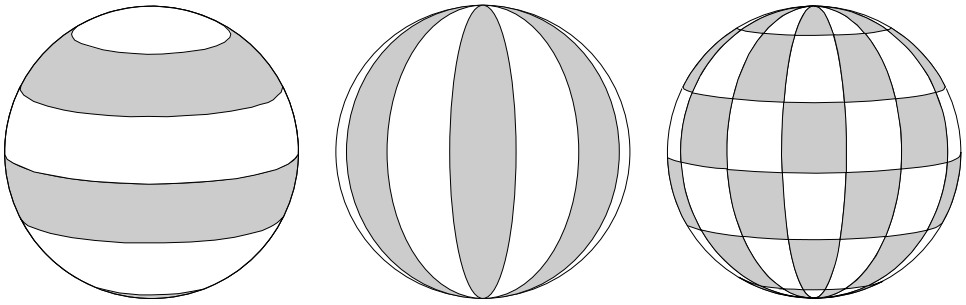
These are the *surface spherical harmonics of degree* n *and order* m. pinta-pallofunktio

Surface spherical harmonics, like also solid spherical harmonics, come in three kinds:

Zonal harmonics $m = 0$. These functions depend only on latitude. vyöhyke-funktiot

Sectorial harmonics $m = n$. the *algebraic signs* of these functions depend only on longitude and not on latitude. The functions themselves however *do* depend on both latitude and longitude! sektorifunktiot

Tesseral harmonics $0 < m < n$. These functions, the algebraic sign of which changes with both latitude and longitude, form a checkerboard pattern on the surface of the sphere, if the positive ruutufunktiot



(a) Zonals: $P_{50}(\sin\phi)$

(b) Sectorials: $P_{66}(\sin\phi)\cos 6\lambda$

(c) Tesserals: $P_{11,6}(\sin\phi)\cos 6\lambda$

FIGURE 3.3. The algebraic signs of spherical harmonics on the Earth's surface. White means positive, grey negative. The functions "wave" in a fashion similar to that of sine or cosine functions.

values are painted white and the negative ones grey (Latin *tessera*: a tile, as used in a mosaic).

Every function will, on the interval $\sin\phi \in [-1, +1]$, go precisely $n - m$ times through zero. Every function is either symmetric or antisymmetric through the origin as a function of ϕ or of $t = \sin\phi$.

Spherical harmonics thus represent a wave phenomenon of sorts. They are however not proper wave functions (sines or cosines): the connection to those is complicated at least. It nevertheless makes sense to speak of their *wavelength*.

pallofunktio

Figure 3.3 depicts how the algebraic signs of the different spherical harmonics behave on the Earth's surface — and above. This is a perspective sketch and not all white and grey areas are visible!

puoli-aallonpituus

In equation 2.11, the expressions $\cos m\lambda$ and $\sin m\lambda$ go around a full circle, the equator, $0° \leqslant \lambda < 360°$, or $0 \leqslant \lambda < 2\pi$, precisely $2m$ times through zero. The "semi-wavelength" is thus

$$\frac{2\pi R}{2m} = \pi\frac{R}{m},$$

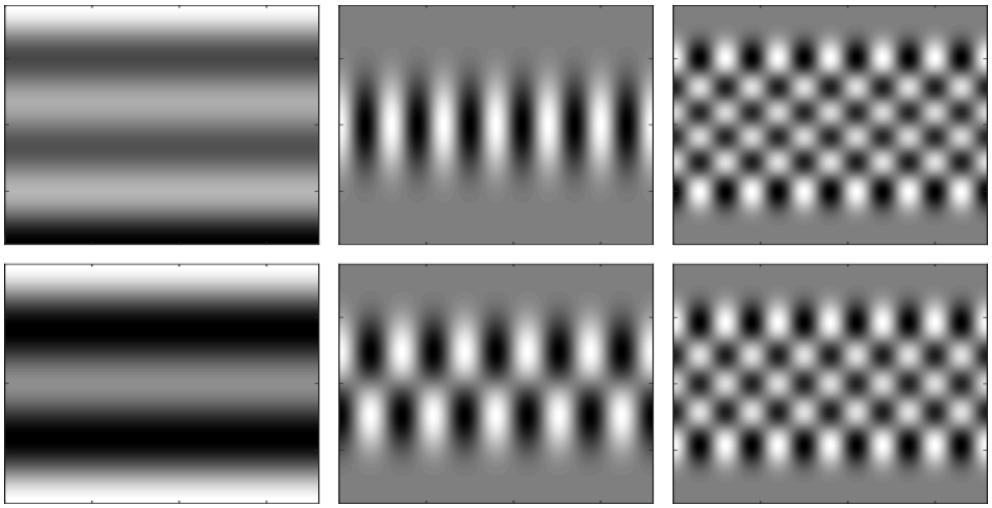


FIGURE 3.4. Surface spherical harmonics as maps. Horizontal axis $\lambda \in [0, 360°) = [0, 2\pi)$, vertical axis $\phi \in [-90°, 90°] = [-\pi/2\,, \pi/2]$. Functions depicted are

$$\begin{matrix} P_{50}(\sin\phi) & P_{66}(\sin\phi)\cos 6\lambda & P_{11,6}(\sin\phi)\cos 6\lambda \\ P_{40}(\sin\phi) & P_{65}(\sin\phi)\cos 5\lambda & P_{10,6}(\sin\phi)\cos 6\lambda \end{matrix}.$$

in which R is again the radius of the Earth.

A similar formula also applies for the functions $P_{nm}(\sin\phi)$: as the function passes through zero $n - m$ times on the interval — from pole to pole — $-90° < \phi < 90°$ or $-\pi/2 < \phi < \pi/2$, it follows that also in this case, a representative semi-wavelength is

$$\frac{\pi R}{n - m}.$$

If we plug various values for m and $n - m$ into this, we obtain table 3.3.

This table also gives the *resolution* that can be achieved with a spherical-harmonic expansion, or in how detailed a fashion the expansion can describe the Earth's gravity field. The expansions available today, like the model EGM2008, go to harmonic degree $n = 2159$; the "sharpness" of a geopotential image based on them is thus 9 km. Models based on satellite orbit perturbations often extend only to degree 20, meaning

TABLE 3.3. Semi-wavelengths for different degrees and orders of spherical harmonics.

m or $n-m$	Semi-wavelength (km)	In degrees
10	2000	18°
40	500	4°.5
180	111	1°
360	55	$30' = 0°.5$
1800	11	$6' = 0°.1$
10 800	1.85	$1' = 0°.017$

pallofunktio-kehitelmä

that only details the size of continents — order of magnitude 1000 km — will be visible. On the other hand, experimental spherical-harmonic expansions of the *topography* go even up to degree 10 800 (Balmino et al., 2012).

3.2 Symmetry properties of the spherical-harmonic expansion

We recapitulate the spherical-harmonic expansion:

$$V(\phi,\lambda,r) = \sum_{n=0}^{\infty} \frac{1}{r^{n+1}} \sum_{m=0}^{n} P_{nm}(\sin\phi)\,(a_{nm}\cos m\lambda + b_{nm}\sin m\lambda)\,. \tag{2.12}$$

3.2.1 *Dependence on latitude* ϕ

It is seen that the dependence on ϕ only works through the Legendre function $P_{nm}(\sin\phi)$. This function can, in terms of mirror symmetry between the northern and southern hemispheres, be either *symmetric* or *antisymmetric* in the argument ϕ. This means that either (symmetric case)

$$P_{nm}(\sin\phi) = P_{nm}\big(\sin(-\phi)\big)$$

or (antisymmetric case)

$$P_{nm}(\sin\phi) = -P_{nm}\big(\sin(-\phi)\big).$$

Equivalently, it means, with $t = \sin\phi$, that either (symmetric case)

$$P_{nm}(t) = P_{nm}(-t)$$

or (antisymmetric case)

$$P_{nm}(t) = -P_{nm}(-t).$$

Which case applies depends on the values of both n and m. To work it out, one can look at, say, equation 3.2:

$$P_{nm}(t) = \left(1-t^2\right)^{m/2} \frac{d^m P_n(t)}{dt^m}. \tag{3.2}$$

We need to answer two questions:

1. For which values n is the polynomial $P_n(t)$ symmetric, for which is it antisymmetric in t? For this, you may look at the recursive algorithm for computation of the polynomials, equation 3.1. We already know that $P_0(t) = 1$ is symmetric, and that $P_1(t) = t$ is antisymmetric. The rule for other n values follows recursively (or you could cheat by looking at table 3.1).
2. What does *differentiation* $\frac{d}{dt}$ do to the symmetry or antisymmetry of the function?

Multiplication by $\sqrt{1-t^2} = \cos\phi$ changes nothing, as this factor is symmetric in t or ϕ.

So, in order to make expansion 2.12 *mirror symmetric* between northern and southern hemispheres, one has to set the coefficients a_{nm}, b_{nm} for which the corresponding P_{nm} is antisymmetric, to zero. Then, those terms vanish from the expansion. The coefficients, and terms, remaining are those for which the corresponding P_{nm} is symmetric.

In tableau 3.4 we give a code fragment in the octave rapid-prototyping language to plot an arbitrary surface spherical harmonic, in order to visually judge its symmetry properties. Do not believe, test.

TABLEAU 3.4. Plotting a surface spherical-harmonic map. Note that the octave code for associated Legendre functions used here contains an additional factor $(-1)^m$, so compared to equation 3.2 all odd orders have the opposite sign.

```
% Plotting surface spherical harmonics
phi=linspace(-90,90,72);
lab=linspace(0,360,144);
[f,l]=meshgrid(phi,lab);
n=5; m=-3;
leg=legendre(n,sin(phi.*pi./180));
if m >= 0
cs=cos(m.*lab.*pi./180);
else
cs=sin(abs(m).*lab.*pi./180);
end
v=leg(abs(m)+1,:)'*cs;
contourf(l,f,v')
xlabel('Longitude', 'FontSize', 16)
ylabel('Latitude', 'FontSize', 16)
str=sprintf('Surface spherical harmonic n=%d, m=%d', n, m)%
title(str, 'FontSize', 20)
axis ([0 360 -90 90])
colorbar()
print('legendre2D.jpg','-djpg')
```

3.2.2 *Dependence on longitude λ*

This dependence works though the "Fourier basis functions" $\cos m\lambda$ and $\sin m\lambda$. The interesting property here is *rotational symmetry*: does the spherical-harmonic expansion 2.12 change when we change the longitude λ?

We see immediately that there will be dependence on λ if any coefficient $a_{nm}, b_{nm}, m \neq 0$ is non-zero. So, in order to obtain rotational symmetry, all coefficients a_{nm} and b_{nm} for values $m > 0$ must be suppressed: $a_{11} = b_{11} = a_{21} = b_{21} = a_{22} = b_{22} = \cdots = 0$.

Of the remaining coefficients, we can say that for $m = 0$, $\sin m\lambda = 0$

identically, so the coefficients b_{00}, b_{10}, b_{20}, ... simply do not matter. They may be any value, including zero. The coefficients a_{00}, a_{10}, a_{20}, ... however *do* matter, as for $m = 0$, $\cos m\lambda = 1$ identically. So we obtain as the *rotationally symmetric expansion*

$$V(\phi, \lambda, r) = V(\phi, r) = \sum_{n=0}^{\infty} \frac{1}{r^{n+1}} a_n P_n(\sin\phi),$$

in which $P_n \stackrel{\text{def}}{=} P_{n0}$ are the familiar Legendre polynomials, and $a_n \stackrel{\text{def}}{=} a_{n0}$.

3.3 Orthogonality of Legendre functions

Legendre's *polynomials* are *orthogonal*: the integral — formally, a scalar product of vectors — is

$$\langle P_n \cdot P_{n'} \rangle_t \stackrel{\text{def}}{=} \int_{-1}^{+1} P_n(t)\, P_{n'}(t)\, dt = \begin{cases} \dfrac{2}{2n+1} & \text{if } n = n', \\ 0 & \text{if } n \neq n'. \end{cases} \tag{3.5}$$

This orthogonality is just one example of a more general way to look at functions and integrals over functions. There exists here a useful analogy with vector spaces, see appendix B.

Alternatively we may write, on the surface of a *unit sphere* σ, using a parametrisation[2] (ψ, α) by angular distance and azimuth, see figure 10.1:

$$\begin{aligned} \iint_\sigma P_n(\cos\psi)\, P_{n'}(\cos\psi)\, d\sigma = \\ = \int_0^{2\pi} \int_0^{\pi} P_n(\cos\psi)\, P_{n'}(\cos\psi) \sin\psi\, d\psi\, d\alpha = \\ = -2\pi \int_{+1}^{-1} P_n(t)\, P_{n'}(t)\, dt = 2\pi \int_{-1}^{+1} P_n(t)\, P_{n'}(t)\, dt, \end{aligned}$$

in which $t = \cos\psi$ and the surface element of the unit sphere $d\sigma = \sin\psi\, d\psi\, d\alpha$, in which again $\sin\psi$ is the *determinant of Jacobi*[3] of the

[2] This parametrisation may be regarded as a latitude, longitude co-ordinate system: the latitude is $90° - \psi = \frac{1}{2}\pi - \psi$, the longitude is α.

co-ordinates (ψ, α). So, we have

$$\langle P_n \cdot P_{n'} \rangle_\sigma \stackrel{\text{def}}{=} \iint_\sigma P_n(\cos\psi)\, P_{n'}(\cos\psi)\, d\sigma = \begin{cases} \dfrac{4\pi}{2n+1} & \text{if } n = n', \\ 0 & \text{if } n \neq n', \end{cases} \tag{3.6}$$

in which ψ is the angular distance from the integration point to the origin $\psi = 0$ of the parametrisation (ψ, α). Equation 3.6 tells us that Legendre polynomials are mutually orthogonal if the vectorial product is defined as an integral over the surface of the unit sphere σ.

Alternatively, we may also define *fully normalised* Legendre polynomials

$$\overline{P}_n(\cos\psi) \stackrel{\text{def}}{=} \sqrt{2n+1} P_n(\cos\psi). \tag{3.7}$$

Now the modified scalar product — the mean product over the unit sphere — is

$$\langle \overline{P}_n \cdot \overline{P}_{n'} \rangle_{\overline{\sigma}} \stackrel{\text{def}}{=} \frac{1}{4\pi} \iint_\sigma \overline{P}_n(\cos\psi)\, \overline{P}_{n'}(\cos\psi)\, d\sigma = \begin{cases} 1 & \text{if } n = n', \\ 0 & \text{if } n \neq n', \end{cases}$$

[4] showing the polynomials now to be *orthonormal*.[4] Similarly fully normalised *associated* Legendre functions also exist, see Heiskanen and Moritz 1967, page 31:

Legendren liitännäisfunktio

$$\overline{P}_{nm}(\cos\psi) \stackrel{\text{def}}{=} \sqrt{2\,(2n+1)\,\frac{(n-m)!}{(n+m)!}} P_{nm}(\cos\psi), \qquad m \neq 0. \tag{3.8}$$

In this case, the orthonormal functions are those of equation 3.3, but

---

[3] Carl Gustav Jacob Jacobi (1804–1851) was a German mathematician known for his work on elliptic functions.

[4] And also

$$\langle \overline{P}_n \cdot \overline{P}_{n'} \rangle_{\overline{t}} \stackrel{\text{def}}{=} \frac{1}{2} \int_{-1}^{+1} \overline{P}_n(t)\, \overline{P}_{n'}(t)\, dt = \begin{cases} 1 & \text{if } n = n', \\ 0 & \text{if } n \neq n', \end{cases}$$

again, the mean product over the domain of integration.

normalized:

$$\overline{Y}_{nm}(\psi,\alpha) = \begin{cases} \overline{P}_{nm}(\cos\psi)\cos m\alpha & \text{if } m \geqslant 0, \\ \overline{P}_{n|m|}(\cos\psi)\sin|m|\,\alpha & \text{if } m < 0. \end{cases}$$

The scalar product that applies is

$$\left\langle \overline{Y}_{nm} \cdot \overline{Y}_{n'm'} \right\rangle_{\overline{\sigma}} \overset{\text{def}}{=} \frac{1}{4\pi} \iint_{\sigma} \overline{Y}_{nm}(\psi,\alpha)\,\overline{Y}_{n'm'}(\psi,\alpha)\,d\sigma = \\ = \begin{cases} 1 & \text{if } n = n' \text{ and } m = m', \\ 0 & \text{otherwise.} \end{cases}$$

3.4 Low-degree spherical harmonics

The potential field of a point mass is (equation 1.4):

$$V = \frac{GM}{r}.$$

The corresponding term in the potential expansion 2.12 for degree $n = 0$ is

$$V_0(\phi,\lambda,r) = \frac{1}{r} a_{00} P_{00}(\sin\phi) = \frac{1}{r} a_{00} P_0(\sin\phi) = \frac{a_{00}}{r},$$

from which

$$a_{00} = GM.$$

So, a_{00} represents the force field of a point mass or spherically symmetric mass distribution centred at the origin. The higher spherical-harmonic coefficients are "perturbations" on top of this.

pallofunktiokerroin

The expansion for the degree-one coefficients looks as follows:

$$V_1(\phi,\lambda,r) = \frac{1}{r^2}\left(a_{11}\cos\phi\cos\lambda + b_{11}\cos\phi\sin\lambda + a_{10}\sin\phi\right).$$

Write this in vector form using the expression for the location vector

$$\mathbf{r} = (r\cos\phi\cos\lambda)\,\mathbf{i} + (r\cos\phi\sin\lambda)\,\mathbf{j} + (r\sin\phi)\,\mathbf{k}$$

— in which $\{\mathbf{i},\mathbf{j},\mathbf{k}\}$ is an orthonormal basis of the Euclidean space —

ortonormaali kanta

yielding

$$V_1(\mathbf{r}) = \frac{1}{r^3}\langle (a_{11}\mathbf{i} + b_{11}\mathbf{j} + a_{10}\mathbf{k}) \cdot \mathbf{r} \rangle.$$

The potential field of a *dipole* is

$$V(\mathbf{r}) = \frac{G}{r^3}\langle \mathbf{d} \cdot \mathbf{r} \rangle,$$

in which $\mathbf{d}$ is the dipole moment. Comparison yields

$$a_{11}\mathbf{i} + b_{11}\mathbf{j} + a_{10}\mathbf{k} = G\mathbf{d},$$

so the first-degree $n = 1$ spherical-harmonic coefficients represent the Earth's gravitational field's *dipole moment*.

Every mass element dm of our Earth may be taken to consist of

- a *monopole* at the origin of the co-ordinate system, magnitude dm
- a *dipole*, magnitude $\mathbf{r}\, dm$, in which $\mathbf{r}$ is the location vector of the mass element.

In that case we may compute the dipole moment of the whole Earth by integration:

$$\mathbf{d}_\oplus = \iiint_\oplus \mathbf{r}\, dm = \iiint_\oplus \rho\mathbf{r}\, d\mathcal{V} = \iiint_\oplus \rho\, d\mathcal{V} \cdot \frac{\iiint_\oplus \rho\mathbf{r}\, d\mathcal{V}}{\iiint_\oplus \rho\, d\mathcal{V}} = M_\oplus \cdot \mathbf{r}_{\text{com}},$$

in which, by definition, $\mathbf{r}_{\text{com}}$ is the location of the *centre of mass* of the Earth. From this follows that, if we choose our co-ordinate system so that the origin is in the centre of mass of the Earth, the spherical-harmonic coefficients a_{11}, b_{11}, and a_{10} *vanish*. If the equations of motion of satellites are formulated in a certain co-ordinate system, like in the case of GPS satellites the WGS84 system, then the origin of the system is automatically in the centre of mass of the Earth, and the degree-one spherical-harmonic coefficients are really zero.

The same logic applies to higher degrees of spherical harmonics. The degree-two coefficients represent the *quadrupole moment* of the Earth — corresponding to her inertial tensor — and so on.

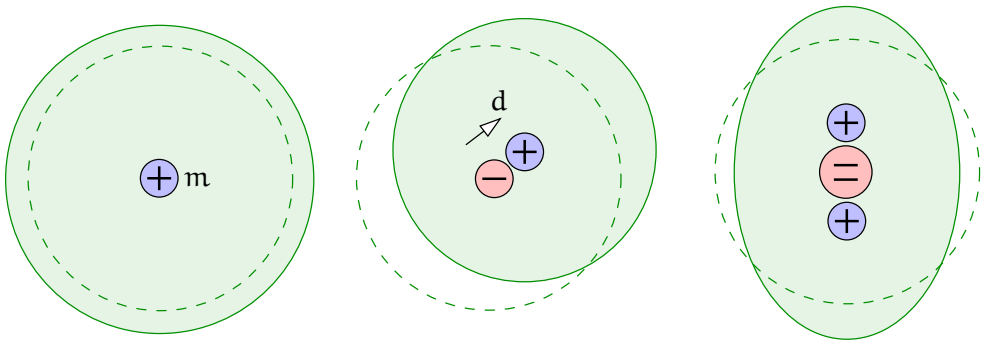


FIGURE 3.5. Monopole, dipole, and quadrupole at the Earth's centre and their effects on the geoid.

3.5 Splitting a function into degree constituents

There exists a useful integral equation for surface spherical harmonics, if the function itself f on the surface of the sphere has been given. The equation is Heiskanen and Moritz (1967) equation 1-71, using our notation $Y_n \to f_n$:

$$f_n(\phi,\lambda) = \frac{2n+1}{4\pi} \iint_\sigma f(\phi',\lambda')\, P_n(\cos\psi)\, d\sigma', \tag{3.9}$$

in which ψ is the geocentric angular distance between evaluation point (ϕ,λ) and moving data or integration point (ϕ',λ'), see figure 8.2. In this *degree constituent equation* 3.9 there is a certain similarity with the projection or coefficient computation equation B.11. Nevertheless, here we do not have a computation of spectral coefficients, but of "spectral constituent functions" f_n.

asteosuus-yhtälö

We bring to mind the core property of the functions f_n,

$$f(\phi,\lambda) = f(\phi,\lambda,R) = \sum_{n=0}^{\infty} f_n(\phi,\lambda)$$

on the surface of the sphere $r = R$.

To prove the degree constituent equation, we choose without loss of generality as the "north pole" of the co-ordinate system the evaluation

point (ϕ, λ), so $\phi = 90°$. Then, $\phi' = 90° - \psi$. By writing (see equation 2.12):

$$f(\phi', \lambda') = \sum_{n=0}^{\infty} \sum_{m=0}^{n} P_{nm}(\cos\psi)\,(a_{nm}\cos m\lambda' + b_{nm}\sin m\lambda'),$$

substituting this into the degree constituent equation 3.9, and exploiting the orthogonality of the Legendre functions, we obtain on the right-hand side:

$$I_R = \frac{2n+1}{4\pi} \iint_\sigma f(\phi', \lambda')\, P_n(\cos\psi)\, d\sigma' = \\ = \frac{2n+1}{4\pi} a_{n0} \iint_\sigma P_n^2(\cos\psi)\, d\sigma'.$$

Then, using equation 3.6:

$$I_R = \frac{2n+1}{4\pi} a_{n0} \frac{4\pi}{2n+1} = a_{n0} = a_n.$$

On the left-hand side of the degree constituent equation we obtain, because on the assumed north pole $\phi = 90°$ and thus $\sin\phi = 1$:

$$I_L = f_n(\phi, \lambda) = f_n(90°, \lambda) = \\ = \sum_{m=0}^{n} P_{nm}(1)\,(a_{nm}\cos m\lambda + b_{nm}\sin m\lambda) = P_{n0}(1)\, a_{n0} = a_n,$$

by using equation 2.11 and

$$P_{n0}(1) = 1,$$
$$P_{nm}(1) = 0, \qquad m \neq 0.$$

As this applies for every point (ϕ, λ), it follows that the degree constituent equation 3.9 is universally true. Note that the values a_n depend on the point choice!

3.6 Spectral representations of various quantities

3.6.1 Potential

Starting from equation 2.10, we write the *spectral expansion* of the geopotential V in space:

$$V(\phi,\lambda,r) = \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+1} V_n(\phi,\lambda), \tag{3.10}$$

in which the degree constituents V_n are

$$V_n(\phi,\lambda) = \frac{Y_n(\phi,\lambda)}{R^{n+1}} = \\ = \frac{1}{R^{n+1}} \sum_{m=0}^{n} P_{nm}(\sin\phi)\,(a_{nm}\cos m\lambda + b_{nm}\sin m\lambda) = \\ = \frac{1}{R^{n+1}} \sum_{m=-n}^{n} \nu_{nm} Y_{nm}(\phi,\lambda).$$

Here, the basis functions Y_{nm} are given by equation 3.3:

$$Y_{nm}(\phi,\lambda) = \begin{cases} P_{nm}(\sin\phi)\cos m\lambda & \text{if } m \geqslant 0, \\ P_{n|m|}(\sin\phi)\sin|m|\lambda & \text{if } m < 0, \end{cases}$$

and the coefficients, equation 3.4:

$$\nu_{nm} = \begin{cases} a_{nm} & \text{if } m \geqslant 0, \\ b_{n|m|} & \text{if } m < 0. \end{cases}$$

On the Earth's surface ($r = R$) this yields

$$V(\phi,\lambda,R) = \sum_{n=0}^{\infty} V_n(\phi,\lambda) = \sum_{n=0}^{\infty} \frac{1}{R^{n+1}} \sum_{m=-n}^{n} \nu_{nm} Y_{nm}(\phi,\lambda). \tag{3.11}$$

We may summarise the relationships found in *commutative diagram* 3.6. Again, as in section 2.2 for rectangular geometry, it is seen that the shift of the potential function V from the spherical level R to the level $r = R + H$ is essentially easier in the frequency domain — the degree constituents $V_n(\phi,\lambda)$ — than it is in the space domain.

kommutoiva kaavio

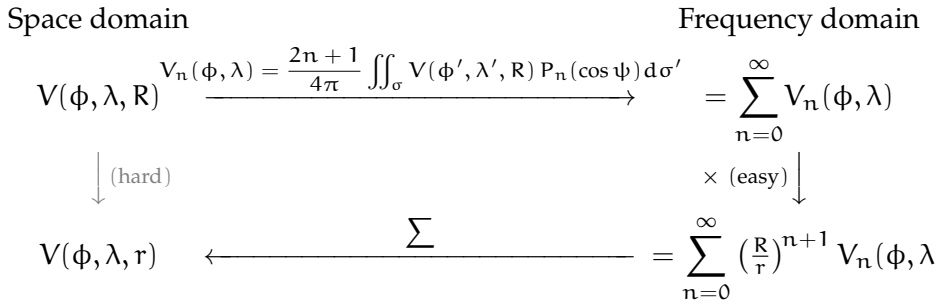


FIGURE 3.6. Vertically or radially shifting the harmonic field V, in the space and frequency (degree number) domains. Spherical geometry.

3.6.2 Gravitation

5 reuna-arvo-tehtävä

In the *Neumann*[5] boundary-value problem we solve a function V of which the normal derivative, $\frac{\partial}{\partial n}V$, is given on a closed surface in space, for example the surface of a body.

If the body is a *sphere*, one may assume $\frac{\partial}{\partial n}V=\frac{\partial}{\partial r}V$ and work with spherical-harmonic expansions. By differentiating equation 3.10 we obtain

$$\frac{\partial V}{\partial r}=-\sum_{n=0}^{\infty}\frac{n+1}{r}\left(\frac{R}{r}\right)^{n+1}V_n(\phi,\lambda)=-\sum_{n=0}^{\infty}\frac{n+1}{R}\left(\frac{R}{r}\right)^{n+2}V_n(\phi,\lambda).$$

At sea level this means

$$\left.\frac{\partial V}{\partial r}\right|_{r=R}=-\sum_{n=0}^{\infty}\frac{n+1}{R}V_n(\phi,\lambda).$$

If we also write at sea level for the *gravitation*

$$g(\phi,\lambda,R)\overset{\text{def}}{=}\left.\frac{\partial V}{\partial r}\right|_{r=R}\overset{\text{def}}{=}\sum_{n=0}^{\infty}g_n(\phi,\lambda),$$

[5] Carl Gottfried Neumann (1832–1925) was a German mathematician who studied the Dirichlet boundary-value problem.

it follows by analogy that

$$g_n(\phi,\lambda) = -\frac{n+1}{R}V_n(\phi,\lambda),$$

and conversely that

$$V_n(\phi,\lambda) = -\frac{R}{n+1}g_n(\phi,\lambda).$$

As a result of this, we obtain the *spectral representation of the solution* to a certain Neumann problem:

$$V(\phi,\lambda,r) = \sum_{n=0}^{\infty}\left(\frac{R}{r}\right)^{n+1}V_n(\phi,\lambda) = -R\sum_{n=0}^{\infty}\left(\frac{R}{r}\right)^{n+1}\frac{g_n(\phi,\lambda)}{n+1}. \quad (3.12)$$

For the gravitation one may write, analogously to expression 3.11 for the potential:

$$g(\phi,\lambda,R) = \sum_{n=0}^{\infty}g_n(\phi,\lambda) \overset{\text{def}}{=} \sum_{n=0}^{\infty}\frac{1}{R^{n+1}}\sum_{m=-n}^{n}g_{nm}Y_{nm}(\phi,\lambda), \quad (3.13)$$

and comparison yields

$$g_{nm} = -\frac{n+1}{R}v_{nm}. \quad (3.14)$$

This is an interesting result worth thinking about:

1. Firstly, note how simple the connection 3.14 between potential v_{nm} and gravitation g_{nm} is in the frequency domain!
2. Secondly, if measurement values of gravitational acceleration $g(\phi,\lambda)$ are available over the whole surface area of the Earth, we may derive from these the degree constituent functions $g_n(\phi,\lambda)$ using the method presented earlier. In this way we can then obtain the solution by means of equation 3.12 for the whole exterior geopotential field! This is the basic idea of geopotential — or geoid — determination, from the spectral perspective.

asteosuus-funktio

3.7 Often-used spherical-harmonic expansions

Of the existing global spherical-harmonic expansions we must mention the already outdated EGM96. It was developed by researchers from Ohio State University using very extensive, mostly gravimetric, data collected by the American NIMA (*National Imagery and Mapping Agency*, the former *Defense Mapping Agency* DMA, the current *National Geospatial-Intelligence Agency* NGA). This expansion goes up to harmonic degree 360. Its standard presentation[6] is

$$V = \frac{GM_\oplus}{r}\left(1 + \sum_{n=2}^{360}\left(\frac{a}{r}\right)^n \sum_{m=0}^{n} \overline{P}_{nm}(\sin\phi)\left(\overline{C}_{nm}\cos m\lambda + \overline{S}_{nm}\sin m\lambda\right)\right). \tag{3.15}$$

This form of presentation — the algebraic sign "+" in front of the expansion, which starts from degree number $n = 2$, the number one inside the parentheses which represents the point mass in the origin equal in magnitude to the total mass of the Earth, and the dimensionless and "fully normalised" coefficients $\overline{C}$ and $\overline{S}$ — is an industry standard in the global research community in the field of computing spherical-harmonic expansions as models of the Earth's gravitational field. Professor Richard H. Rapp at Ohio State University has been a pioneer, which is why the models are often called OSU models.

ratahäiriöt

Generally in these models the lower terms — $2 \leqslant n \leqslant 20$ — are derived primarily from analysis of satellite orbit perturbations. Because of this, the models are in a co-ordinate system with the origin in the Earth's centre of mass. This explains the absence of the degree-one coefficients, as explained earlier.

The higher coefficients again — $20 < n \leqslant 360$ — were before the year 2000 mostly the result of the analysis of gravimetric data (over land)

[6] Here $a = a_\oplus$ is used to signify the equatorial radius of the Earth's reference ellipsoid, not R, and ϕ signifies *geocentric* latitude. The co-ordinates (ϕ, λ, r) form a spherical co-ordinate system.

and satellite radar altimetric data (over the ocean). After the launches of the gravimetric satellite missions CHAMP, GRACE, and GOCE, and as a result of their measurements, nowadays at least the harmonic degree interval $20 < n \leqslant 200$ is the product of space geodesy. Only the still higher-degree coefficients continue to come from terrestrial data. The newer model EGM2008 (Pavlis et al., 2012) goes up to degree 2159.

In tableau 3.5 we give the first and last coefficients of the EGM96 model, the newest and best spherical-harmonics model from the time just before the satellite gravity missions. The values tabulated are n, m, $\overline{C}_{nm}$, $\overline{S}_{nm}$ and the mean errors (standard deviations) of both coefficients from their computation. Note that all $\overline{S}_{n0}$ vanish!

pallofunktio-malli

Sometimes non-normalised coefficients are also used, and we write

$$V = \frac{GM_\oplus}{r}\left(1 - \sum_{n=2}^{\infty}\left(\frac{a}{r}\right)^n \sum_{m=0}^{n} P_{nm}(\sin\phi)\,(J_{nm}\cos m\lambda + K_{nm}\sin m\lambda)\right). \tag{3.16}$$

Then we use the notation $J_n \overset{\text{def}}{=} J_{n0}$. The coefficient J_2 is the most important spherical-harmonic coefficient of the Earth's gravity field, expressing the flattening of the Earth. Based on equations 3.7 and 3.8, the relationship with the parameters $\overline{C}, \overline{S}$ is

$$\begin{aligned}
\left\{\begin{array}{c} J_{n0} \\ K_{n0} \end{array}\right\} &= -\sqrt{2n+1}\left\{\begin{array}{c} \overline{C}_{n0} \\ \overline{S}_{n0} \end{array}\right\}, \\
\left\{\begin{array}{c} J_{nm} \\ K_{nm} \end{array}\right\} &= -\sqrt{2\,(2n+1)\,\frac{(n-m)!}{(n+m)!}}\left\{\begin{array}{c} \overline{C}_{nm} \\ \overline{S}_{nm} \end{array}\right\}, \qquad m \neq 0.
\end{aligned} \tag{3.17}$$

3.8 Ellipsoidal harmonics

The Laplace differential equation 1.13 may be written and solved in ellipsoidal co-ordinates instead of spherical co-ordinates. The result is known as an *ellipsoidal-harmonic expansion*.[7] They are little-used,

ellipsoidi-funktio-kehitelmä

[7] This expansion for the ellipsoid of revolution differs from the expansion into Lamé functions found for the triaxial ellipsoid.

TABLEAU 3.5. Coefficients and mean errors of the EGM96 spherical-harmonic expansion.

n	m	$\overline{C}_{nm}$	$\overline{S}_{nm}$	$\overline{C}_{nm}$ mean error	$\overline{S}_{nm}$ mean error
2	0	−0.484165371736E−03	0.000000000000E+00	0.35610635E−10	0.00000000E+00
2	1	−0.186987635955E−09	0.119528012031E−08	0.10000000E−29	0.10000000E−29
2	2	0.243914352398E−05	−0.140016683654E−05	0.53739154E−10	0.54353269E−10
3	0	0.957254173792E−06	0.000000000000E+00	0.18094237E−10	0.00000000E+00
3	1	0.904627768605E−06	0.248513158716E−06	0.13965165E−09	0.13645882E−09
3	2	0.904627768605E−06	−0.619025944205E−06	0.10962329E−09	0.11182866E−09
3	3	0.721072657057E−06	0.141435626958E−05	0.95156281E−10	0.93285090E−10
4	0	0.539873863789E−06	0.000000000000E+00	0.10423678E−09	0.00000000E+00
4	1	−0.536321616971E−06	−0.473440265853E−06	0.85674404E−10	0.82408489E−10
4	2	0.350694105785E−06	0.662671572540E−06	0.16000186E−09	0.16390576E−09
4	3	0.990771803829E−06	−0.200928369177E−06	0.84657802E−10	0.82662506E−10
4	4	−0.188560802735E−06	0.308853169333E−06	0.87315359E−10	0.87852819E−10
5	0	0.685323475630E−07	0.000000000000E+00	0.54383090E−10	0.00000000E+00
5	1	−0.621012128528E−07	−0.944226127525E−07	0.27996887E−09	0.28082882E−09
5	2	0.652438297612E−06	−0.323349612668E−06	0.23747375E−09	0.24356998E−09
5	3	−0.451955406071E−06	−0.214847190624E−06	0.17111636E−09	0.16810647E−09
5	4	−0.295301647654E−06	0.496658876769E−07	0.11981266E−09	0.11849793E−09
5	5	0.174971983203E−06	−0.669384278219E−06	0.11642563E−09	0.11590031E−09
6	0	−0.149957994714E−06	0.000000000000E+00	0.14497863E−09	0.00000000E+00
6	1	−0.760879384947E−07	0.262890545501E−07	0.22415138E−09	0.21957296E−09
6	2	0.481732442832E−07	−0.373728201347E−06	0.27697363E−09	0.28105811E−09
6	3	0.571730990516E−07	0.902694517163E−08	0.19432407E−09	0.18682712E−09
6	4	−0.862142660109E−07	−0.471408154267E−06	0.15229150E−09	0.15328004E−09
6	5	−0.267133325490E−06	−0.536488432483E−06	0.89838470E−10	0.87820905E−10
6	6	0.967616121092E−08	−0.237192006935E−06	0.11332010E−09	0.11518036E−09
⋮	⋮				
360	358	0.709604781531E−10	0.691761006753E−10	0.50033977E−10	0.50033977E−10
360	359	0.183971631467E−10	−0.310123632209E−10	0.50033977E−10	0.50033977E−10
360	360	−0.447516389678E−24	−0.830224945525E−10	0.50033977E−10	0.50033977E−10

because the maths needed is more complicated. Moreover, ellipsoidal co-ordinates are mostly only theoretically interesting and not in any broad use within geodesy.

The form of presentation is

$$V(\beta, \lambda, u) =$$

TABLE 3.6. Legendre functions of the second kind.

$Q_0(z) = \frac{1}{2}\ln\frac{z+1}{z-1}$	$(n+1)Q_{n+1}(z) - (2n+1)\,zQ_n(z) + nQ_{n-1}(z) = 0$
$Q_1(z) = \frac{1}{2}z\ln\frac{z+1}{z-1} - 1$	
$Q_2(z) = (\frac{3}{4}z^2 - \frac{1}{4})\ln\frac{z+1}{z-1} - \frac{3}{2}z$	$Q_{nm}(z) = (1-z^2)^{m/2}\,\frac{d^m}{dz^m}Q_n(z)$
$Q_3(z) = (\frac{5}{4}z^3 - \frac{3}{4}z)\ln\frac{z+1}{z-1} - \frac{5}{2}z^2 + \frac{2}{3}$	

$$= \sum_{n=0}^{\infty}\sum_{m=0}^{n}\frac{Q_{nm}(i\frac{u}{E})}{Q_{nm}(i\frac{b}{E})}P_{nm}(\sin\beta)\,(A^e_{nm}\cos m\lambda + B^e_{nm}\sin m\lambda)\,, \quad (3.18)$$

in which $Q_{nm}(z)$ are the *Legendre functions of the second kind*, sampled in table 3.6. Although the general argument z is complex, equation 3.18 gives a real result for real-valued coefficients A^e_{nm}, B^e_{nm}.

Those interested in the derivation of the above equation 3.18 can find it in Heiskanen and Moritz (1967) section 1-20 or other textbooks on potential theory.

3.8.1 The scaling to standard form of the expansion

Assume the origin to be in the centre of mass of the Earth, so $A^e_{10} \approx 0$, $A^e_{11} \approx 0$, $B^e_{11} \approx 0$.

We can also show that in expansion 3.18 the first coefficient must be

$$A^e_{00} = A^e_0 = \frac{GM_\oplus}{E}\arctan\frac{E}{b}$$

and the expansion specialised for a rotationally symmetric field becomes

$$V^*(\beta,u) = \sum_{n=0}^{\infty}V^*_n(\beta,u) = \sum_{n=0}^{\infty}\frac{Q_n(i\frac{u}{E})}{Q_n(i\frac{b}{E})}A^{e*}_{n0}\,P_n(\sin\beta). \quad (3.19)$$

Also

$$V_0(u) = V^*_0(u) = \frac{Q_0(i\frac{u}{E})}{Q_0(i\frac{b}{E})}\frac{GM_\oplus}{E}\arctan\frac{E}{b},$$

the gravitational potential of the field constituent of ellipsoidal degree zero.

With the substitutions (Heiskanen and Moritz, 1967, page 66)

$$Q_0\left(i\frac{u}{E}\right) = -i\arctan\frac{E}{u}, \qquad Q_0\left(i\frac{b}{E}\right) = -i\arctan\frac{E}{b} \tag{3.20}$$

we obtain

$$V_0(u) = V_0^*(u) = \frac{GM_\oplus}{E}\arctan\frac{E}{u}. \tag{3.21}$$

This corresponds to the "central field" of a spherical-harmonic expansion $GM_\oplus/r$. Using this, we may *scale* equation 3.18 by substituting the above identities 3.20. The coefficients need to be divided by the constant expression

$$\frac{GM_\oplus}{E}\arctan\frac{E}{b},$$

as the central field, expression 3.21, is moved outside the expansion. The result is

$$V(\beta,\lambda,u) = \frac{GM_\oplus}{E}\arctan\frac{E}{u}\cdot \\ \cdot\left(1+\sum_{n=2}^{\infty}\sum_{m=0}^{n}\frac{\arctan\frac{E}{b}}{\arctan\frac{E}{u}}\frac{Q_{nm}\left(i\frac{u}{E}\right)}{Q_{nm}\left(i\frac{b}{E}\right)}\overline{P}_{nm}(\sin\beta)\left(\overline{C}^{e}_{nm}\cos m\lambda+\overline{S}^{e}_{nm}\sin m\lambda\right)\right),$$

in which we have also introduced fully normalized coefficients $\overline{C}^{e}_{nm}$, $\overline{S}^{e}_{nm}$ and Legendre functions $\overline{P}_{nm}(\sin\beta)$.

This is an ellipsoidal-harmonic expansion that agrees with the standard-form spherical-harmonic expansion 3.15, with the total mass of the Earth outside the parentheses and the coefficients dimensionless. This equation has apparently not been used for any geopotential determination.

3.8.2 *Equivalence of the Rapp and ellipsoidal expansions*

We can demonstrate the equivalence of spherical expansions 3.15 or 3.16 and ellipsoidal expansion 3.18, if the flattening of the Earth $\to 0$,

and thus also $b \to a$, $\beta \to \phi$, and $u \to r$. We assume that Heiskanen and Moritz (1967) equation 1-112,

$$\lim_{E\to 0} \frac{Q_{nm}\left(i\frac{u}{E}\right)}{Q_{nm}\left(i\frac{b}{E}\right)} = \left(\frac{a}{r}\right)^{n+1}$$

is valid. Substitution into equation 3.18 yields

$$\begin{aligned} &V(u,\beta,\lambda) = V(r,\phi,\lambda) = \\ &= \sum_{n=0}^{\infty}\sum_{m=0}^{n} \left(\frac{a}{r}\right)^{n+1} P_{nm}(\sin\phi)\left(A^e_{nm}\cos m\lambda + B^e_{nm}\sin m\lambda\right), \end{aligned} \quad (3.22)$$

which, with the identifications $A^e_{00} = GM_\oplus/a$, $A^e_{10} = A^e_{11} = B^e_{11} = 0$ and relations 3.17, suggests

$$\begin{Bmatrix} A^e_{n0} \\ B^e_{n0} \end{Bmatrix} = -\frac{GM_\oplus}{a}\begin{Bmatrix} J_{n0} \\ K_{n0} \end{Bmatrix} = \frac{GM_\oplus}{a}\sqrt{(2n+1)}\begin{Bmatrix} \overline{C}_{n0} \\ \overline{S}_{n0} \end{Bmatrix},$$

$$\begin{Bmatrix} A^e_{nm} \\ B^e_{nm} \end{Bmatrix} = -\frac{GM_\oplus}{a}\begin{Bmatrix} J_{nm} \\ K_{nm} \end{Bmatrix} =$$

$$= \frac{GM_\oplus}{a}\sqrt{2\,(2n+1)\,\frac{(n-m)!}{(n+m)!}}\begin{Bmatrix} \overline{C}_{nm} \\ \overline{S}_{nm} \end{Bmatrix}, \qquad m \neq 0.$$

Substituting these into equation 3.22 affirms its equivalence with equations 3.15 and 3.16 for spherical harmonics.

3.8.3 *Advantages of using ellipsoidal harmonics*

- The expression for the normal gravitational potential is simple in this form of presentation, see Heiskanen and Moritz (1967) equation 2-56. A spherical-harmonic expansion of the same field would instead require theoretically an infinite number of coefficients. In practice, this number is only 3 to 4, so an expansion up to J_6 or J_8 will suffice.
- The convergence behaviour on a flattened Earth will be better. suppeneminen

This is because, due to the Earth's flattening, the equator is some 21 km further from the Earth's centre than the poles. Therefore, high-degree spherical harmonics in particular will have difficulty converging efficiently both at the poles and in the equatorial region. This problem is worst for very high-degree expansions (for example Wenzel, 1998). Already for a degree number of 360, the semi-wavelength of a spherical harmonic will be only 55 km!

puoli-aallonpituus

3.8.4 *Disadvantage of using ellipsoidal harmonics*

Evaluation of an ellipsoidal-harmonic expansion is clearly more laborious and expensive than that of a spherical-harmonic one, in terms of computer resources.

Self-test questions

1. What are the harmonic degree and harmonic order in a spherical-harmonic expansion? How do they relate to the *resolution* of the expansion on the Earth's surface?
2. What types of spherical harmonics are there? Explain their dependence on latitude and longitude.
3. How many times does a surface spherical harmonic $Y_{nm}(\phi, \lambda)$ change its algebraic sign travelling along a meridian from the south pole to the north pole? How many times when travelling around the Earth along the equator?
4. What does it mean if it is said that two functions are mutually orthogonal? Give a possible definition of the *scalar product* of two functions.
5. How does the attenuation of spherical harmonics with height behave? Why does a gravimetric satellite that is trying to map the gravity field of the Earth at a high resolution fly in as low an orbit as possible?

6. What does the degree constituent equation express?
7. Which spherical-harmonic coefficients are associated with the dipole moment of the Earth's mass distribution? Why are they missing from tableau 3.5?

Exercise 3–1: Attenuation with height of a spherical-harmonic expansion

If

$$V(\phi,\lambda,r) = \sum_{n=0}^{\infty} V_n(\phi,\lambda,r) = \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+1} V_n(\phi,\lambda),$$

we may call

$$\frac{V_n(\phi,\lambda,r)}{V_n(\phi,\lambda)} = \left(\frac{R}{r}\right)^{n+1}$$

the *attenuation factor* of the potential with height.

Differentiation with respect to r yields

$$\frac{\partial V_n(\phi,\lambda,r)}{\partial r} = -\frac{n+1}{R}\left(\frac{R}{r}\right)^{n+2} V_n(\phi,\lambda), \tag{3.23}$$

or, because, at sea level, similarly

$$\left.\frac{\partial V_n(\phi,\lambda,r)}{\partial r}\right|_{r=R} = -\frac{n+1}{R} V_n(\phi,\lambda), \tag{3.24}$$

it follows that the attenuation factor for the *attraction* is the ratio of expressions 3.23 and 3.24:

$$\left(\frac{R}{r}\right)^{n+2}.$$

1. Draw a *log-linear graph* of the attenuation factors of both the potential and the attraction for values $n = 0, 1, 2, \ldots, 100$, by hand or by machine. Choose $R = 6378\,\text{km}$, $r = 7378\,\text{km}$ — a height 1000 km above the Earth's surface.
2. Based on this, if the satellite is 1000 km above the Earth's surface, for what degree number n will the accelerations $\frac{\partial}{\partial r} V_n(\phi,\lambda,r)$ caused by the attraction at the satellite's level be less than 1 % of what they are on the Earth's surface?

3. For what degree number n will they be less than $10^{-4}\times$ of what they are on the Earth's surface?

Exercise 3–2: Symmetries of spherical harmonics

See equation 2.12. In it, $P_{nm}(\sin\phi) = P_{nm}(t)$ *is only a function of latitude* ϕ. When ϕ runs from the south pole through the equator to the north pole, $-90° \leqslant \phi \leqslant +90°$, parameter t will run through the values $-1 \leqslant t \leqslant +1$.

For the Legendre functions exists the closed expression 3.2:

$$P_{nm}(t) = \left(1-t^2\right)^{m/2} \frac{d^m}{dt^m} P_n(t),$$

in which the $P_n(t)$ are ordinary Legendre polynomials:

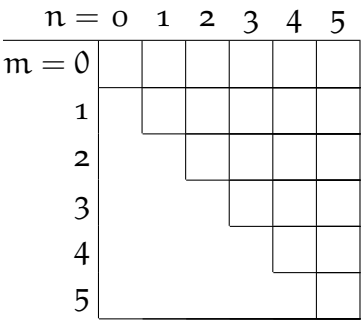
$$P_n(t) = \frac{1}{2^n n!} \frac{d^n}{dt^n} \left(t^2-1\right)^n.$$

We can observe the following properties:

- Differentiating a symmetric function of t will produce an antisymmetric function, and vice versa.
- The function $\left(t^2-1\right)$ and its powers are symmetric.
- Thus: for even n values, $P_n(t) = P_n(-t)$: P_n is *symmetric* between the northern and southern hemispheres, and for odd n values $P_n(t) = -P_n(-t)$: P_n is *antisymmetric* between hemispheres.
- Similarly, for even n, $P_n(\sin\phi) = P_n\big(\sin(-\phi)\big)$, and for odd n, $P_n(\sin\phi) = -P_n\big(\sin(-\phi)\big)$.

Questions

1. What is the corresponding rule for the functions P_{nm}, in other words, for which values n and m is it symmetric and for which values antisymmetric?
2. Fill in the diagram ($n = 0, \ldots, 5$, $m = 0, \ldots, n$) with either 'S' (symmetric) or 'A' (antisymmetric) in each framed cell:



3. What is the logic of symmetry?
4. If the field is *mirror symmetric* between the northern and southern hemispheres, so $V(\phi, \lambda, r) = V(-\phi, \lambda, r)$, which of the spherical-harmonic coefficients a_{nm} and b_{nm} drop out of the series expansion? *Why*?

   Hint: see the example formulas and graphs for $P_{nm}(\sin\phi)$ in this chapter and try to guess a general rule. Then, verify.
5. The same question if the potential is *rotationally symmetric* about the Earth's rotation axis: $V(\phi, \lambda, r) = V(\phi, r)$.

Exercise 3–3: Algebraic-sign domains of spherical harmonics

We have seen in section 3.1 that the associated Legendre functions $P_{nm}(t)$ have precisely $n - m + 1$ algebraic-sign intervals on their interval of definition $\phi \in [-90°, 90°]$. We can show that the functions $\cos m\lambda$ and $\sin m\lambda$ each have $2m$ zero crossings and $2m$ algebraic-sign intervals on their domain of definition $\lambda \in [0, 360°)$, assumed to form a closed circle. How many *algebraic-sign domains* — grey or white areas, visible or occluded — are there in figure 3.3 for each surface spherical harmonic

$$Y_{nm}(\phi, \lambda) = \begin{cases} P_{nm}(\sin\phi)\cos m\lambda & \text{if } m \geqslant 0, \\ P_{n|m|}(\sin\phi)\sin|m|\lambda & \text{if } m < 0 \end{cases} \quad ?$$

Exercise 3−4: Escape velocity

1. Given a spherically symmetric planet, mass GM, radius R, from the surface of which a cannon shoots projectiles at flight velocity v. What is the minimum value for v — the *escape velocity* — if it is desired that the projectile can travel to arbitrarily large distances from the planet and never fall back? The kinetic energy of the projectile is $E_{kin} = \frac{1}{2}mv^2$, in which m is the projectile's mass.
2. Given, in *two-dimensional geometry*, a circularly symmetric planet, mass GM, radius R, with the gravitational field of the planet represented by potential V as given in section 2.3, what does V look like in terms of those parameters? Make an educated guess.
3. There is again a cannon on the edge of the circle planet. What can you now say about the escape velocity v (do *not* try to compute it!)?

The normal gravity field

4

4.1 The basic idea of a normal field

Just as the figure of the Earth can be approximated by an ellipsoid of revolution, the gravity field of the Earth can also be approximated by a field of which one equipotential surface, or *level surface*, is precisely this ellipsoid of revolution, the *reference ellipsoid*.

pyörähdys-ellipsoidi

vertaus-ellipsoidi

This brings a logical idea to mind: why not define intercompatibly a reference ellipsoid and a model geopotential or *normal potential*, one of the equipotential surfaces of which is the reference ellipsoid? After that, a *gravity formula* is obtained by taking the gradient of this normal potential.

After this we may define *anomalous* quantities, such as the disturbing potential and the gravity anomaly, which then again will be intercompatible, while being numerically much smaller.

häiriö-potentiaali

Let the normal potential be $U(x, y, z)$. Then, normal gravity will be

$$\gamma\,(x, y, z) = \|\boldsymbol{\gamma}\| = \|\nabla U\| = -\langle \boldsymbol{\gamma} \cdot \mathbf{n} \rangle = -\frac{\partial U}{\partial n},$$

in which $\frac{\partial}{\partial n}$ denotes differentiation in the direction of the exterior surface normal $\mathbf{n}$ to a level surface of the normal field, itself an ellipsoid as well, see figure 4.1. This direction will differ from the direction of the normal to the level surfaces of the gravity field, or *plumb line*, by an

luotiviiva

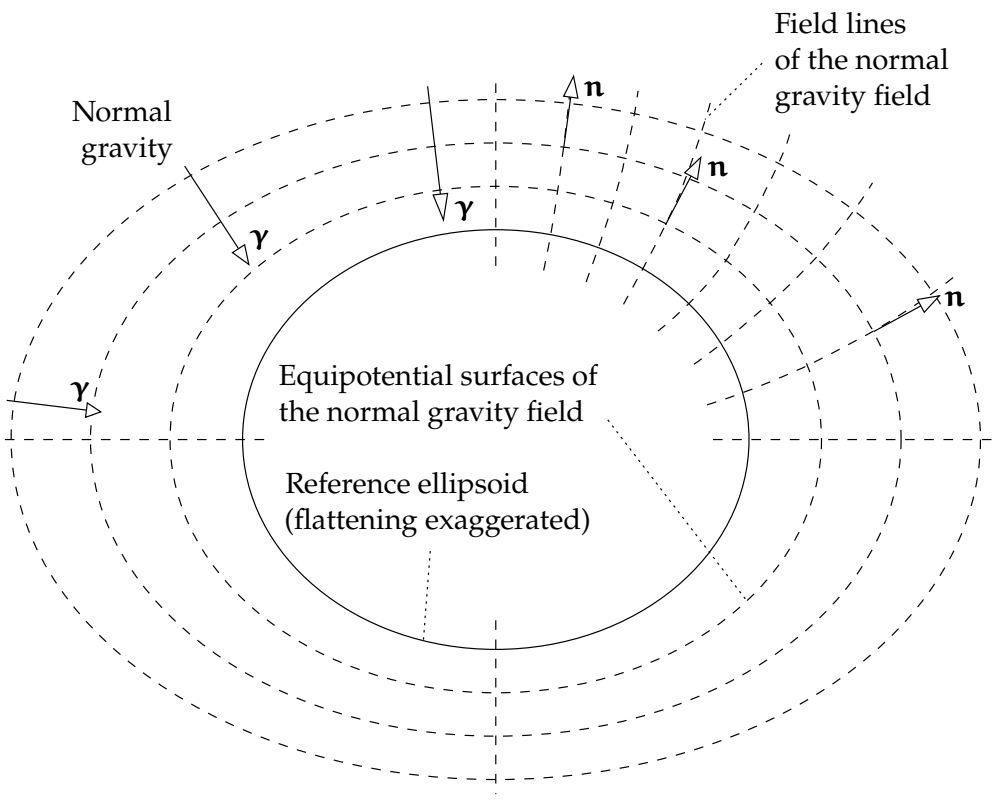


FIGURE 4.1. The normal gravity field of the Earth.

amount called the *plumb-line deflection*. This deflection of the plumb line is also a very small angle.

luotiviivan poikkeama

We shall see in the next section that the pseudo-force generated by the Earth's rotation may, in a system rotating along with the Earth, be described by a *rotational potential* Φ — also called centrifugal potential. The normal potential U is also defined in such a way that the rotational potential Φ is included in it: the normal potential is the reference potential of the *gravity field*, not the *gravitational field*. If we denote the normal *gravitational* potential by V^* — a quantity rarely used in geodesy — then the normal *gravity* potential or *normal potential* U is

vertaus-potentiaali

$$U = V^* + \Phi,$$

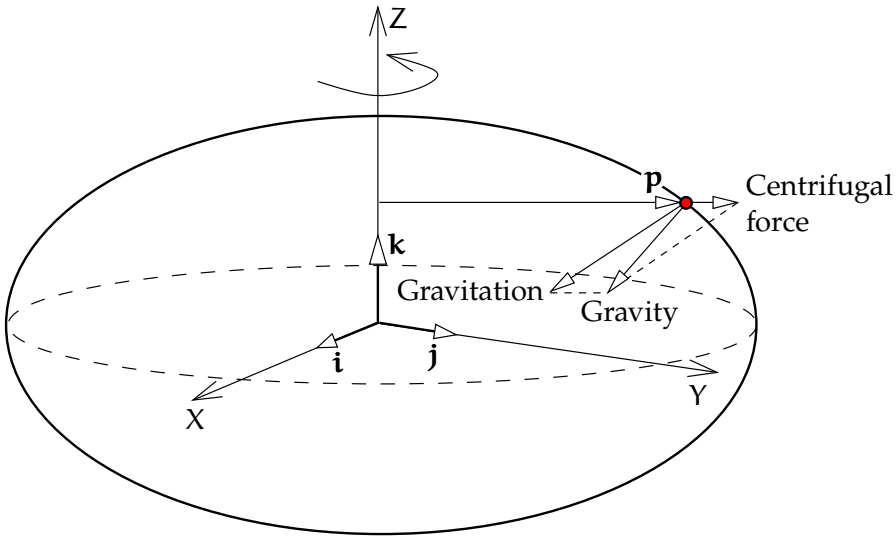


FIGURE 4.2. Gravitation and centrifugal force.

in which Φ is the centrifugal potential. In other words: V^*, like V, is defined in a non-rotating or inertial system, whereas U, like W, is defined in a system that co-rotates with the Earth and is non-inertial. The word *gravity* refers to a force acting in a co-rotating system, whereas in an inertial system we use the word *gravitation* or attraction.

4.2 The centrifugal force and its potential

The rotation of the Earth affects the gravity field. In an inertial reference system one speaks of gravitation and gravitational potential V. On the Earth's surface, however, in a non-inertial or *co-rotating* system, we talk of *gravity* and *gravity potential* W. They are different things, and the rotational motion and its centrifugal force are the cause of the difference. See figure 4.2.

vertaus-järjestelmä

To derive the equation for centrifugal force, write first

$$\mathbf{p} = X\mathbf{i} + Y\mathbf{j}.$$

The vectors $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ form an orthonormal basis along the (X, Y, Z) axes.

ortonormaali kanta

It follows that

$$p = \|\mathbf{p}\| = \sqrt{\langle \mathbf{p} \cdot \mathbf{p} \rangle} = \sqrt{X^2 + Y^2}.$$

Now the centrifugal force — or rather, acceleration — is

$$\mathbf{f}_\omega = \omega_\oplus^2 \, \mathbf{p} = \omega_\oplus^2 \, (X\mathbf{i} + Y\mathbf{j}) \,, \tag{4.1}$$

with $\omega_\oplus$ the rotation rate of the Earth in radians per unit of time. If X and Y are in metres and $\omega_\oplus$ is in radians per second, then $\mathbf{f}_\omega$ is obtained in $^{m}/_{s^2}$.

Here on Earth, gravity measurements are generally made using an instrument that is *at rest* with respect to the Earth's surface: it follows the rotation of the Earth. If the instrument moves, one must, in addition to the centrifugal force, take into account another pseudo-force: the Coriolis[1] force. Fluids — water, air — on the Earth's surface, if they are at rest, only sense gravity, which includes the centrifugal force. Currents in addition also sense the Coriolis force, which deflects them sideways and causes the well-known vortex phenomena in the oceans and atmosphere, such as hurricanes.

We may describe centrifugal force as the gradient of a potential. If we write for this *centrifugal potential*

$$\Phi = \tfrac{1}{2}\omega_\oplus^2 \left(X^2 + Y^2\right) ,$$

we may directly calculate the gradient

$$\begin{aligned} \mathbf{f}_\omega = \nabla\Phi = \frac{\partial\Phi}{\partial X}\mathbf{i} + \frac{\partial\Phi}{\partial Y}\mathbf{j} + \frac{\partial\Phi}{\partial Z}\mathbf{k} = \\ = \tfrac{1}{2}\omega_\oplus^2 \cdot 2X \cdot \mathbf{i} + \tfrac{1}{2}\omega_\oplus^2 \cdot 2Y \cdot \mathbf{j} + 0 = \omega_\oplus^2 \, (X\mathbf{i} + Y\mathbf{j}) \,, \end{aligned}$$

which corresponds to the above centrifugal-force equation 4.1.

[1] Gaspard-Gustave Coriolis (1792–1843) was a French mathematician, physicist and mechanical engineer. His name is inscribed on the Eiffel Tower, Eiffel Tower, 72 names.

If we add to the gravitational potential V the centrifugal potential Φ, we obtain the *gravity potential* or *geopotential* W:

$$W = V + \Phi.$$

We may also derive from the centrifugal potential Φ the following equation by differentiating it twice:

$$\Delta\Phi = \nabla^2\Phi = \langle \nabla \cdot \mathbf{f}_\omega \rangle = \frac{\partial}{\partial X}\omega_\oplus^2 X + \frac{\partial}{\partial Y}\omega_\oplus^2 Y + \frac{\partial}{\partial Z}0 = 2\omega_\oplus^2, \qquad (4.2)$$

from which follows, with Poisson equation 1.14,

$$\Delta W = -4\pi G\rho + 2\omega_\oplus^2, \qquad (4.3)$$

the Poisson equation for the geopotential or gravity potential.

The difference between gravitation and gravity is essential. The force, or acceleration, of gravitation $\mathbf{g}^* = \nabla V$ is just an attractive force, whereas the acceleration of gravity $\mathbf{g} = \nabla W$ is the vector sum of gravitation and centrifugal force. Attraction and centrifugal force act in the same fashion: the force is proportional to the mass of the test object. In other words, the acceleration is always the same independently of the mass of the test object. This is the famous *equivalence principle* (Galileo, Einstein), which has been proven to hold to very great precision. We may mention in particular the clever tests by the Hungarian Loránd Eötvös.[2]

Water masses on the Earth's surface, as also the atmosphere — and on a vastly longer time-scale also the "solid" Earth rock forming mountain ranges and ocean depths — react to gravity without distinguishing between attraction and centrifugal force. For this reason, the sea surface coincides within a metre or so with an equipotential or *level surface* of the geopotential W. Moreover, on dry land, we measure heights from this surface, the *geoid* — according to Gauss, the "mathematical figure of the Earth".

[2]Loránd baron Eötvös de Vásárosnamény (1848–1919) was a Hungarian physicist and student of gravitation.

4.3 Level surfaces and plumb lines

Surfaces of constant gravity potential or *geopotential*, equipotential surfaces or *level surfaces*, are the following surfaces:

$$W(x, y, z) = \text{constant}.$$

Let $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ be an orthonormal basis along the (x, y, z) axes. Then, in the direction of the unit vector

$$\boldsymbol{e} = e_1\mathbf{i} + e_2\mathbf{j} + e_3\mathbf{k}$$

the potential changes as follows:

$$\frac{\partial W}{\partial e} = e_1\frac{\partial W}{\partial x} + e_2\frac{\partial W}{\partial y} + e_3\frac{\partial W}{\partial z} = \langle \boldsymbol{e} \cdot \nabla W \rangle ,$$

which vanishes if and only if the vectors $\boldsymbol{e}$ and ∇W are perpendicular to each other. In other words, the potential is stationary only in directions that are perpendicular to the Earth's gravity vector

$$\nabla W = \mathbf{g}.$$

Level surfaces and gravity vectors, or plumb lines, are always perpendicular to each other.

4.3.1 Curvature of level surfaces

The plane that at P has the same direction as the level surface is called its *tangent plane*, figure 4.3. If the local curvature of the level surface in the x direction is ρ_x and the x co-ordinate of point P is x_0, we may develop the distance between the surfaces in a Taylor series:

$$\epsilon \approx \frac{1}{2\rho_x}(x - x_0)^2 .$$

From this follows the difference in W values between the surfaces ($g = \|\mathbf{g}\| = \|\nabla W\|$):

$$\delta W \approx -\epsilon g \approx -(x - x_0)^2 \frac{g}{2\rho_x}.$$

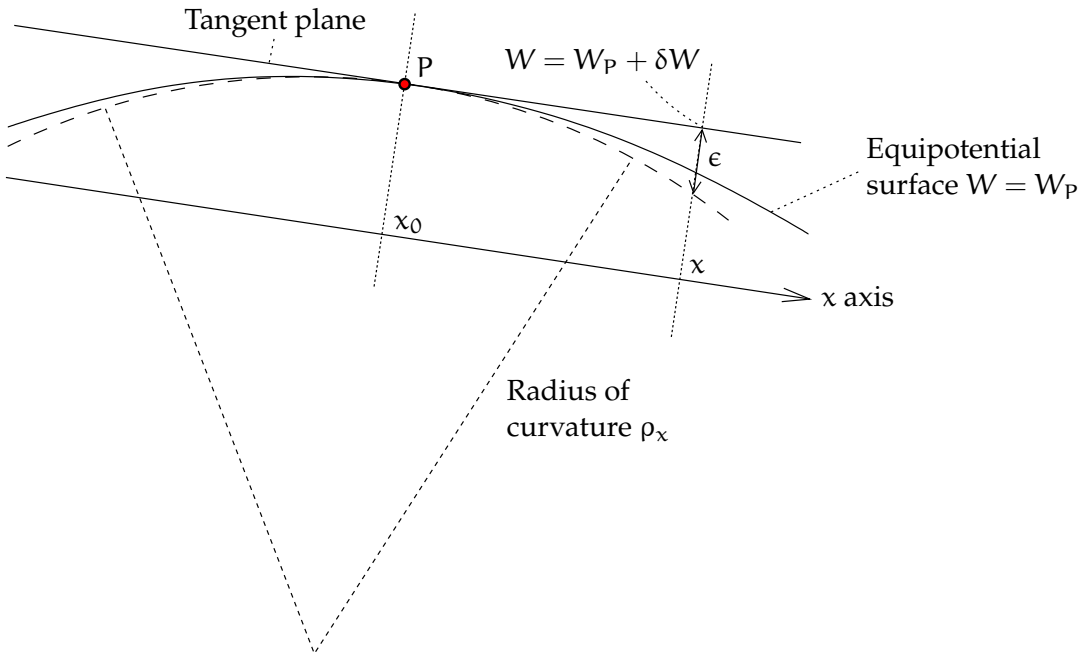


FIGURE 4.3. The curvature of level surfaces.

Differentiating (note that W here is now the geopotential *on the tangent or horizontal plane*), yields[3]

$$\frac{\partial^2}{\partial x^2}\delta W = \frac{\partial^2}{\partial x^2}W = \partial_{xx}W = -\frac{g}{\rho_x},$$

from which

$$\rho_x = -\frac{g}{\partial_{xx}W}.$$

By determining the curvature in the x and y directions,

$$K_x \overset{\text{def}}{=} \frac{1}{\rho_x} = -\frac{\partial_{xx}W}{g}, \quad K_y \overset{\text{def}}{=} \frac{1}{\rho_y} = -\frac{\partial_{yy}W}{g}, \tag{4.4}$$

we obtain the mean or Germain[4] curvature, in most locations a positive number:

$$J = \frac{K_x + K_y}{2} = -\frac{\partial_{xx}W + \partial_{yy}W}{2g},$$

[3] We use here compact *Euler notation* for partial derivatives, ∂_{xx}, ∂_{yy}, ∂_{zz}, which is often convenient.

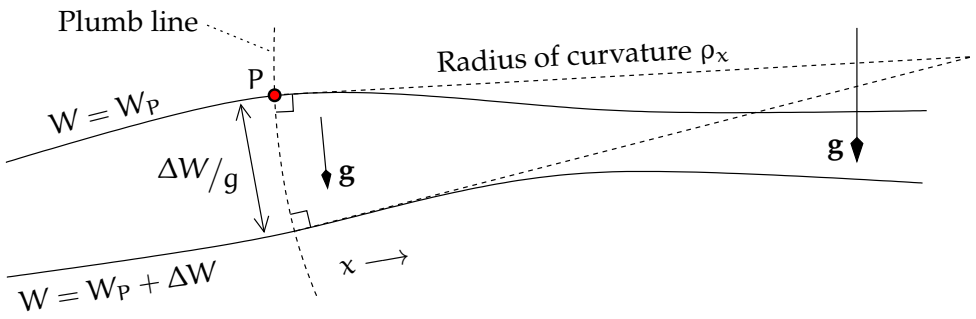


FIGURE 4.4. The curvature of the plumb line.

and by using Poisson equation 4.3,

$$\Delta W = \partial_{xx}W + \partial_{yy}W + \partial_{zz}W = -4\pi G\rho + 2\omega_\oplus^2,$$

we obtain

$$-2gJ + \partial_{zz}W = -4\pi G\rho + 2\omega_\oplus^2.$$

By using

$$\partial_{zz}W = -\frac{\partial g}{\partial z} = -\frac{\partial g}{\partial H},$$

in which H is the height co-ordinate, we obtain for the vertical gradient of gravity (Heiskanen and Moritz, 1967, equation 2-20):

$$\frac{\partial g}{\partial H} = -2gJ + 4\pi G\rho - 2\omega_\oplus^2,$$

an equation found by Ernst Heinrich Bruns (Bruns, 1878, page 13).

4.3.2 *Curvature of plumb lines*

luotiviiva Plumb lines are curved because gravity is not constant in the horizontal direction. If gravity increases in a horizontal direction, then the equipotential surfaces will come closer together too, and they will not be

[4]Marie-Sophie Germain (1776–1831) was a brilliant French mathematician, number theorist and student of elasticity. She corresponded with Gauss on number theory (Friedelmeyer, 2014) and did foundational work towards a proof of Fermat's last theorem. Her name is missing from the Eiffel Tower.

parallel. This means that the plumb lines, being perpendicular to all equipotential surfaces, must be curved in that direction.

Consider two equipotential surfaces, one for potential W_P and one for potential $W_P + \Delta W$. The distance separating them will be $\Delta H = {}^{\Delta W}/_{g}$. In the direction of co-ordinate x the relative tilt between the two surfaces will be

$$\frac{\partial}{\partial x}\Delta H(x) = \frac{\partial}{\partial x}\left(\frac{\Delta W}{g(x)}\right) = -\frac{\Delta W}{g^2}\frac{\partial g}{\partial x}.$$

If the starting distance between the surfaces is ΔH, it will take a distance of

$$\rho_x = -\Delta H \Big/ \frac{\partial}{\partial x}\Delta H = -\left(\frac{\Delta W}{g}\right) \Big/ \left(-\frac{\Delta W}{g^2}\frac{\partial g}{\partial x}\right) = g \Big/ \frac{\partial g}{\partial x}$$

to bring the tangents together, see figure 4.4. The curvature of the plumb line is the inverse of this, in both the x and the y co-ordinate directions:

$$\kappa_x = \frac{1}{\rho_x} = \frac{1}{g}\frac{\partial g}{\partial x}, \quad \kappa_y = \frac{1}{\rho_y} = \frac{1}{g}\frac{\partial g}{\partial y}.$$

We can derive the curvature of the field lines of the normal gravity field, or *normal plumb lines*, in the same way. The difference is, however, that we can find a simple mathematical expression for gravity on the surface of the reference ellipsoid, for example equation 4.8. A good approximation is

$$\gamma(\varphi) \approx \gamma_a \cos^2\varphi + \gamma_b \sin^2\varphi.$$

With the chain rule

$$\frac{\partial\gamma}{\partial x} = \frac{\partial\gamma}{\partial\varphi}\frac{\partial\varphi}{\partial x} = \frac{1}{R}\frac{\partial\gamma}{\partial\varphi} = \frac{1}{R}\left(-2\gamma_a \cos\varphi\sin\varphi + 2\gamma_b \sin\varphi\cos\varphi\right) = \\ = \frac{\gamma_b - \gamma_a}{R}\sin 2\varphi.$$

This means in the x or south-north and y or west-east direction:

$$\kappa^*_x = \frac{1}{\gamma}\frac{\partial\gamma}{\partial x} \approx \frac{1}{R}\frac{\gamma_b - \gamma_a}{\gamma_a}\sin 2\varphi, \quad \kappa^*_y = \frac{1}{\gamma}\frac{\partial\gamma}{\partial y} = 0.$$

This also means that the direction of the normal plumb line at height h is $\varphi(h) = \varphi(0) + \kappa^*_x h$, in which numerically $\kappa^*_x = 0.171'' \,\mathrm{km}^{-1} \cdot \sin 2\varphi$ (Heiskanen and Moritz, 1967, equation 5-34).

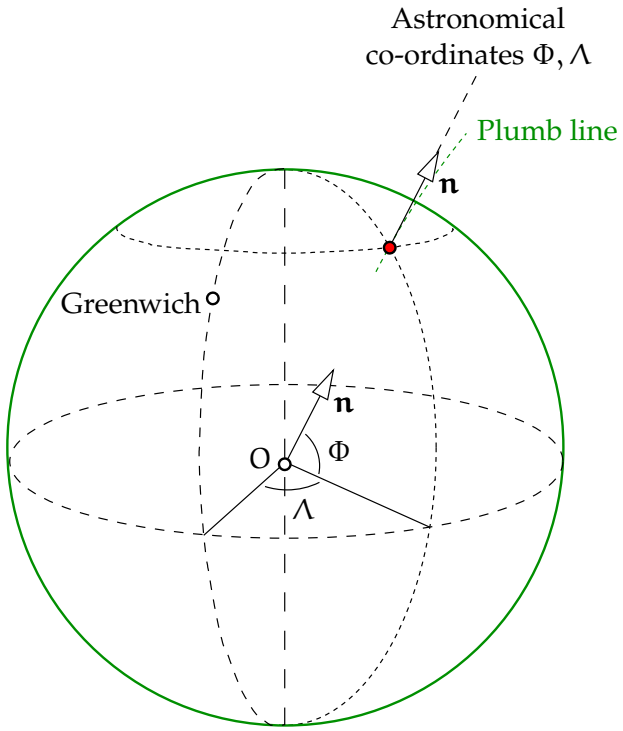


FIGURE 4.5. Natural co-ordinates Φ and Λ. In addition, a natural height co-ordinate, for example the geopotential W, is needed.

4.4 Natural co-ordinates

Before the satellite era it was impossible to directly measure the geocentric co-ordinates X, Y, and Z. Today this is possible, and we obtain at the same time the height h from the reference ellipsoid, a purely geometric quantity.

In earlier times, one could measure only the direction of the plumb line as shown in figure 4.5, as well as the potential difference between an observation point and sea level by levelling. The direction of the plumb line $\mathbf{n}$ was measured astronomically: astronomical latitude Φ and astronomical longitude Λ. The third co-ordinate, the gravity potential difference $W(x, y, z) - W_0$ from the potential W_0 of sea level, was determined by levelling. Co-ordinates Φ, Λ, and W are called

natural co-ordinates.

Instead of the potential, *orthometric height* H may be used. Its definition is easy to understand if one writes

$$\frac{\partial W}{\partial H} = -g \implies dH = -\frac{1}{g}dW \implies H_P = -\int_{W_0}^{W_P} \frac{1}{g(W)}dW, \quad (4.5)$$

in which the integral is taken along the plumb line of point P. $\frac{\partial}{\partial H}$ is the derivative in the direction of the plumb line, the local normal to the level surfaces. g is the acceleration of gravity along the plumb line as a function of place — or of geopotential level. In this case of orthometric heights, g is the *true* gravity inside the rock, which is a non-linear function of place and will also depend on the rock density. This trickiness of their determination is a problem specific to orthometric heights. We will return to this later (Heiskanen and Moritz, 1967, chapter 4).

The co-ordinates Φ, Λ, and H also form a natural co-ordinate system.

4.5 The normal potential in ellipsoidal co-ordinates

We have already presented equation 3.18, the expansion of the geopotential into ellipsoidal harmonics. The normal potential U is required to be constant on the reference ellipsoid $u = b$. We expand the centrifugal potential Φ into ellipsoidal harmonics, obtaining

ellipsoidi-
funktio

$$\begin{aligned}\Phi(\beta, u) = \tfrac{1}{2}\omega_\oplus^2\left(x^2 + y^2\right) = \tfrac{1}{2}\omega_\oplus^2\left(u^2 + E^2\right)\cos^2\beta = \\ = \tfrac{1}{2}\omega_\oplus^2\left(u^2 + E^2\right)\left(1 - \sin^2\beta\right) = \\ = \tfrac{1}{2}\omega_\oplus^2\left(u^2 + E^2\right)\left(-\tfrac{2}{3}P_2(\sin\beta) + \tfrac{2}{3}P_0(\sin\beta)\right) = \\ = -\tfrac{1}{3}\omega_\oplus^2\left(u^2 + E^2\right)\left(P_2(\sin\beta) - P_0(\sin\beta)\right).\end{aligned}$$

In addition, based on equation 3.19 we have for the rotationally symmetric normal gravitational potential V^*:

$$V^*(\beta, u) = \sum_{n=0}^{\infty} V_n^*(\beta, u) = \sum_{n=0}^{\infty} \frac{Q_n\left(i\frac{u}{E}\right)}{Q_n\left(i\frac{b}{E}\right)} A_{n0}^{e*} P_n(\sin\beta). \quad (3.19)$$

Now

$$U(\beta, u) = V^*(\beta, u) + \Phi(\beta, u).$$

On the reference ellipsoid $u = b$ we have as a requirement $U(\beta, b) = U_0$, which is possible only if ($A_n^{e*} \stackrel{\text{def}}{=} A_{n0}^{e*}$):

$$\begin{aligned} U_0 &= A_0^{e*} + \tfrac{1}{3}\omega_\oplus^2 \left(b^2 + E^2\right) = A_0^{e*} + \tfrac{1}{3}\omega_\oplus^2 a^2, \\ 0 &= A_1^{e*}, \\ 0 &= A_2^{e*} - \tfrac{1}{3}\omega_\oplus^2 \left(b^2 + E^2\right) = A_2^{e*} - \tfrac{1}{3}\omega_\oplus^2 a^2, \\ 0 &= A_n^{e*}, \qquad n = 3, 4, 5, \ldots. \end{aligned}$$

The quantity U_0 can be computed uniquely, if the Earth's mass $GM_\oplus$ and the measures of the reference ellipsoid a and b are known. The result, Heiskanen and Moritz (1967) equation 2-61, is

$$U_0 = \frac{GM_\oplus}{E} \arctan\frac{E}{b} + \tfrac{1}{3}\omega_\oplus^2 a^2. \tag{4.6}$$

From this follows

$$A_0^{e*} = U_0 - \tfrac{1}{3}\omega_\oplus^2 a^2 = \frac{GM_\oplus}{E} \arctan\frac{E}{b}.$$

The normal *gravity* potential U is obtained as follows:

$$\begin{aligned} U(\beta, u) = V^*(\beta, u) + \Phi(\beta, u) &= \overbrace{\frac{GM_\oplus}{E} \arctan\frac{E}{u}}^{V_0^*(u)} + \\ + \overbrace{\tfrac{1}{3}\omega_\oplus^2 a^2}^{A_2^{e*}} \frac{Q_2\left(i\frac{u}{E}\right)}{Q_2\left(i\frac{b}{E}\right)} &\overbrace{\left(\tfrac{3}{2}\sin^2\beta - \tfrac{1}{2}\right)}^{P_2(\sin\beta)} + \overbrace{\tfrac{1}{2}\omega_\oplus^2 \left(u^2 + E^2\right)\cos^2\beta}^{\Phi(\beta, u)} = \\ &= C_0(u) + C_1(u)\sin^2\beta + C_2(u)\cos^2\beta, \end{aligned}$$

in which C_0, C_1, and C_2 are suitable functions of u. The function V_0^* is the term for $n = 0$ in expansion 3.19, equation 3.21.

On the surface of the reference ellipsoid ($u = b$), using $a^2 = b^2 + E^2$:

$$U(\beta, b) =$$

$$= \overbrace{\frac{GM_\oplus}{E}\arctan\frac{E}{b}}^{V_0^*(b)} + \overbrace{\tfrac{1}{2}\omega_\oplus^2 a^2\sin^2\beta - \tfrac{1}{6}\omega_\oplus^2 a^2}^{A_2^{e*}\,P_2(\sin\beta)} + \overbrace{\tfrac{1}{2}\omega_\oplus^2 a^2\cos^2\beta}^{\Phi(\beta,b)} =$$
$$= \frac{GM_\oplus}{E}\arctan\frac{E}{b} + \tfrac{1}{3}\omega_\oplus^2 a^2,$$

the constant U_0 (equation 4.6), as it had better be!

4.6 Normal gravity on the reference ellipsoid

Without proof, we mention that for *normal gravity* (the quantity $\gamma = -\frac{\partial}{\partial h}U$) the following equation applies on the reference ellipsoid:

$$\gamma(\beta) = \frac{a\gamma_b\sin^2\beta + b\gamma_a\cos^2\beta}{\sqrt{a^2\sin^2\beta + b^2\cos^2\beta}}. \tag{4.7}$$

It is seen that γ_a is normal gravity on the equator ($\beta = 0$) and γ_b normal gravity on the poles ($\beta = \pm 90°$).

Equations 2.5 and 2.8 yield

$$\tan\beta = \frac{\sin\beta}{\cos\beta} = \frac{Z/b}{\sqrt{X^2+Y^2}/a} = \frac{a}{b}\frac{Z}{\sqrt{X^2+Y^2}} = \frac{a}{b}\tan\phi$$

and

$$\tan\varphi = \frac{\sin\varphi}{\cos\varphi} = \frac{Z\Big/(1-e^2)\,N}{\sqrt{X^2+Y^2}/N} = \frac{1}{1-e^2}\frac{Z}{\sqrt{X^2+Y^2}} = \frac{a^2}{b^2}\tan\phi,$$

in which ϕ is the *geocentric* latitude, see equations 2.4. From this follows directly

$$\tan\beta = \frac{b}{a}\tan\varphi,$$

in which the latitude angle φ is the geodetic or geographic latitude. β is still the *reduced latitude*. Now it can be shown (exercise!) that

redukoitu leveysaste

$$\gamma(\varphi) = \frac{a\gamma_a\cos^2\varphi + b\gamma_b\sin^2\varphi}{\sqrt{a^2\cos^2\varphi + b^2\sin^2\varphi}}. \tag{4.8}$$

This is the famous *Somigliana–Pizzetti*[5] equation. These geodesists de- [5]

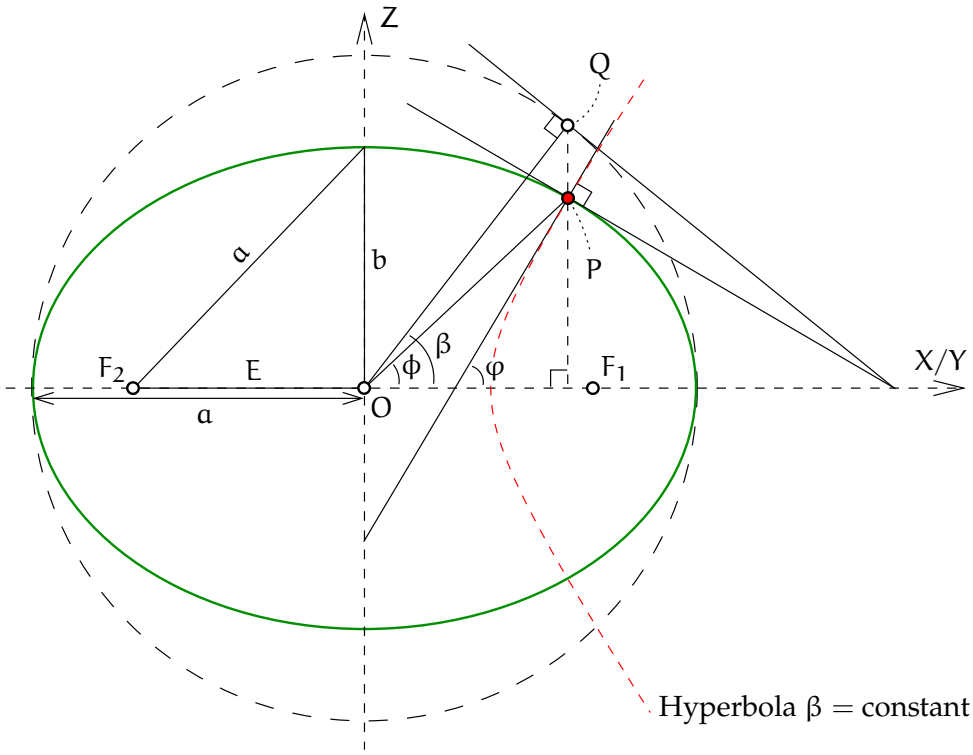


FIGURE 4.6. Geometry of the meridian ellipse and various types of latitude, as well as the focal points F_1 and F_2.

monstrated for the first time that an "ellipsoidal" normal gravity field which has the reference ellipsoid as one of its equipotential or level surfaces exists exactly, and that the gravity formula is also a closed expression in geographic latitude.

4.7 Numerical values and calculation formulas

When the reference ellipsoid has been chosen, we may calculate the normal potential and normal gravity corresponding to it. The fundamental quantities are

[5] Carlo Somigliana (1860–1955) was an Italian mathematician and physicist. Paolo Pizzetti (1860–1918) was an Italian geodesist.

a the equatorial radius of the ellipsoid of revolution, its semimajor axis — isoakselin puolikas

f the flattening

$$f \stackrel{\text{def}}{=} \frac{a-b}{a},$$

in which b is the polar radius or semiminor axis

$\omega_\oplus$ the rotation rate of the Earth

$GM_\oplus$ the total mass of the Earth, including the atmosphere.

Nowadays the most commonly used reference ellipsoid *cum* normal potential is the GRS80, the Geodetic Reference System 1980:

$$a = 6\,378\,137\,\text{m}, \qquad \omega_\oplus = 7\,292\,115 \cdot 10^{-11}\,\text{s}^{-1},$$
$$\frac{1}{f} = 298.257\,222\,101, \qquad GM_\oplus = 3\,986\,005 \cdot 10^{8}\,\text{m}^3/\text{s}^2.$$

In reality, f is not a defining constant of the GRS80, but the constant J_2 is used instead, which is a defining quantity for the gravitational field, see equation 3.16.

The WGS84 (World Geodetic System 1984) used by the GPS has a reference ellipsoid that is *almost* identical to that of the GRS80.

The normal potential is (Heikkinen, 1981), in SI units:

$$\begin{aligned} U \approx\ & 62\,636\,860.8500 + \\ & + \left(\begin{array}{c} -9.780\,326\,77 - 0.051\,630\,75 \sin^2\varphi - \\ -0.000\,227\,61 \sin^4\varphi - 0.000\,001\,23 \sin^6\varphi \end{array} \right) h + \\ & + \left(\begin{array}{c} +0.015\,438\,99 \cdot 10^{-4} - 0.000\,021\,95 \cdot 10^{-4} \sin^2\varphi - \\ -0.000\,000\,10 \cdot 10^{-4} \sin^4\varphi \end{array} \right) h^2 + \\ & + \left(-0.000\,024\,22 \cdot 10^{-8} + 0.000\,000\,07 \cdot 10^{-8} \sin^2\varphi \right) h^3, \end{aligned} \quad (4.9)$$

and normal gravity (note the minus sign, U is positive and diminishes going upwards):

$$\gamma = -\frac{\partial U}{\partial h} \approx +9.780\,326\,77 + 0.051\,630\,75 \sin^2\varphi +$$

$$+ 0.000\,227\,61 \sin^4\varphi + 0.000\,001\,23 \sin^6\varphi + \\ - \left(\begin{array}{c} + 0.030\,877\,98 \cdot 10^{-4} - 0.000\,043\,90 \cdot 10^{-4} \sin^2\varphi - \\ - 0.000\,000\,20 \cdot 10^{-4} \sin^4\varphi \end{array} \right) h + \\ - \left(- 0.000\,072\,65 \cdot 10^{-8} + 0.000\,000\,21 \cdot 10^{-8} \sin^2\varphi \right) h^2. \quad (4.10)$$

Here, the unit of potential is $^{m^2}/_{s^2}$, and the unit of gravity, $^{m}/_{s^2}$. φ is geodetic latitude; h (in metres) is the height above the reference ellipsoid. More precise equations can be found from Heikkinen (1981). In these equations, the coefficient $9.780\,32\ldots$ $^{m}/_{s^2}$ is equatorial gravity, and the value $-0.030\,87\ldots \cdot 10^{-4}\ s^{-2}$ is the vertical gradient of gravity on the equator.

Other gravity formulas and reference ellipsoids still in legacy use (and slowly fading away) are Helmert's 1906 ellipsoid, the Krasovsky ellipsoid or SK-42 in Eastern Europe, the International or Hayford ellipsoid (1924) and its gravity formula, and the Geodetic Reference System 1967.

4.7.1 Numerical example

According to equation 4.9, the normal potential over the equator is

$$U = 62\,636\,860.8500 - 9.780\,326\,77\,h + 0.015\,438\,99 \cdot 10^{-4}\,h^2 - \\ - 0.000\,024\,22 \cdot 10^{-8}\,h^3.$$

- Draw this function for values of h in the range 0–7000 km.
- Draw for comparison the quadratic version, from which the last term is left off.

Questions

1. What is the minimum of the quadratic function?
2. How physically realistic is this?

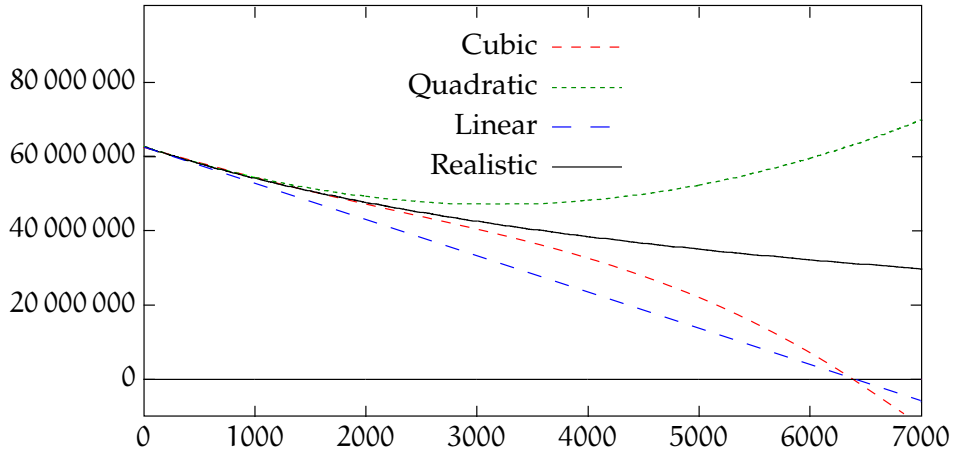


FIGURE 4.7. The normal field's potential over the equator. Heights in kilometres, potential in m^2/s^2.

Answers

1. See figure 4.7. The minimum of the quadratic function is at height 3000 km. The cubic function does not have a minimum.
2. Not very realistic: the stationary point for potential U (the normal potential in a co-rotating system) should be located at approximately 36 000 km height, at the geostationary orbit.

This tells us that polynomial approximation cannot be extrapolated very far. In this case, the interval of extrapolation is of the same order as the radius of the Earth, and that will no longer work.

4.8 The normal potential as a spherical-harmonic expansion

pallofunktio-kehitelmä

The spherical-harmonic expansion of an ellipsoidal gravitational field also contains, besides the second-degree harmonic, higher-degree harmonics. If we write, as is customary, the potential outside the Earth

TABLE 4.1. GRS80 normal potential spherical-harmonic coefficients (Heikkinen, 1981; Heiskanen and Moritz, 1967).

Non-normalised	Fully normalised
$J_2^* = J_{2,0}^* = 1082.63 \cdot 10^{-6}$	$\overline{J}_2^* = -\overline{C}_{2,0}^* = 484.166\,854\,896 \cdot 10^{-6}$
$J_4^* = J_{4,0}^* = -2.370\,912\,22 \cdot 10^{-6}$	$\overline{J}_4^* = -\overline{C}_{4,0}^* = -0.790\,304\,073 \cdot 10^{-6}$
$J_6^* = J_{6,0}^* = +0.006\,083\,47 \cdot 10^{-6}$	$\overline{J}_6^* = -\overline{C}_{6,0}^* = +0.001\,687\,251 \cdot 10^{-6}$
$J_8^* = J_{8,0}^* = -0.000\,014\,27 \cdot 10^{-6}$	$\overline{J}_8^* = -\overline{C}_{8,0}^* = -0.000\,003\,461 \cdot 10^{-6}$

in the following form (Heiskanen and Moritz, 1967 equation 2-39, also equation 3.16):

$$V(\phi,\lambda,r) = \frac{GM_\oplus}{r}\left(1 - \sum_{n=2}^{\infty}\left(\frac{a}{r}\right)^n \sum_{m=0}^{n} P_{nm}(\sin\phi)\left(J_{nm}\cos m\lambda + K_{nm}\sin m\lambda\right)\right),$$

then we may also write the normal gravitational potential, V^*, into the form

$$V^*(\phi,r) = \frac{GM_\oplus}{r}\left(1 - \sum_{\substack{n=2\\ \text{even}}}^{\infty} J_n^*\left(\frac{a}{r}\right)^n P_n(\sin\phi)\right),$$

which contains only even coefficients $J_n^* \overset{\text{def}}{=} J_{n0}^*$, because the normal field is *symmetric* about the equatorial plane.

The coefficients for the GRS80 normal gravitational potential are found[6] in table 4.1. Higher terms are usually not needed. The relationship between fully normalised and non-normalised coefficients is $J_n^* = \overline{J}_n^*\sqrt{2n+1}$.

For comparison: in section 4.5 it was shown that in the expansion of the same field into *ellipsoidal* harmonics, only the degree-zero and

[6] They can also be calculated from equation 2-92 given in Heiskanen and Moritz (1967):

$$J_{2n}^* = (-1)^{n+1}\frac{3\left(e^2\right)^n}{(2n+1)(2n+3)}\left(1 - n + 5n\frac{J_2}{e^2}\right),$$

starting from the values J_2 and e^2. The results are the same as in the table's left column.

degree-two coefficients are non-zero! This is one reason why these functions are used at all.

Instead of using an ellipsoidal model, we may also use as a normal gravity potential formula the first two or three terms of the spherical-harmonic expansion of the real geopotential. Then we obtain, taking the centrifugal potential along:

$$U = \frac{Y_0}{r} + \frac{Y_2(\phi,\lambda)}{r^3} + \tfrac{1}{2}\omega_\oplus^2\left(X^2+Y^2\right),$$

with the corresponding equipotential surface $U = U_0$ being the "Bruns spheroid", or

$$U = \frac{Y_0}{r} + \frac{Y_2(\phi,\lambda)}{r^3} + \frac{Y_4(\phi,\lambda)}{r^5} + \tfrac{1}{2}\omega_\oplus^2\left(X^2+Y^2\right),$$

the "Helmert spheroid". Here, $Y_0 \stackrel{\text{def}}{=} GM_\oplus$ while $Y_2(\phi,\lambda)$ and $Y_4(\phi,\lambda)$ are taken from the true geopotential.

These equations are easy to compute, but their equipotential or level surfaces are not ellipsoids of revolution, and in fact not even rotationally symmetric. They are quite complicated surfaces (Heiskanen and Moritz, 1967, section 2-12)!

However, in geometric geodesy we always use a reference ellipsoid, so this is also a wise thing to do in physical geodesy.

4.9 The disturbing potential

Write the gravity potential

$$W = V + \Phi,$$

in which Φ is the centrifugal potential (see above), and the normal potential

$$U = V^* + \Phi.$$

The difference between them is the *disturbing potential* häiriö-potentiaali

$$T \stackrel{\text{def}}{=} W - U = V - V^*.$$

pallofunktio Both V and V^* can be expanded into spherical harmonics. If we write the gravity potential

$$W = V + \Phi = \Phi + \\ + \frac{GM_\oplus}{r}\left(1 - \sum_{n=2}^{\infty}\left(\frac{a}{r}\right)^n \sum_{m=0}^{n} P_{nm}(\sin\phi)\left(J_{nm}\cos m\lambda + K_{nm}\sin m\lambda\right)\right),$$

and the normal potential

$$U = \Phi + \frac{GM_\oplus}{r}\left(1 - \sum_{\substack{n=2\\ \text{even}}}^{\infty}\left(\frac{a}{r}\right)^n J_n^* P_n(\sin\phi)\right),$$

we obtain by subtraction for the disturbing potential

$$T = W - U = -\frac{GM_\oplus}{r}\left(\sum_{n=2}^{\infty}\left(\frac{a}{r}\right)^n \sum_{m=0}^{n} P_{nm}(\sin\phi)\left(\delta J_{nm}\cos m\lambda + K_{nm}\sin m\lambda\right)\right), \tag{4.11}$$

in which

$$\begin{cases} \delta J_{n0} = J_{n0} - J_n^* & \text{if } n \text{ even},\\ \delta J_{nm} = J_{nm} & \text{otherwise.}\end{cases}$$

The above equation for the disturbing potential T is shortened as follows (Heiskanen and Moritz, 1967, equation 2-152):

$$T(\phi,\lambda,r) = \sum_{n=2}^{\infty}\left(\frac{a}{r}\right)^{n+1} T_n(\phi,\lambda), \tag{4.12}$$

asteosuus where, in every term, the degree constituent T_n has the same dimension as T, and

$$T_n(\phi,\lambda) = -\frac{GM_\oplus}{a}\sum_{m=0}^{n} P_{nm}(\sin\phi)\left(\delta J_{nm}\cos m\lambda + K_{nm}\sin m\lambda\right).$$

vertauspallo On the surface of the reference sphere of radius a:[7]

7

$$T(\phi,\lambda) = T(\phi,\lambda,a) = \sum_{n=2}^{\infty} T_n(\phi,\lambda),$$

from which we see that on the reference level, the terms $T_n(\phi,\lambda)$ are really the *degree constituents* of the disturbing potential T for a certain degree number n.

The above expansions are all missing the terms $n = 0, 1$. Of these, $T_0(\phi,\lambda) = T_0$ is a constant — the global average of the disturbing potential — and $T_1(\phi,\lambda)$ has the form of a dipole field. Its value is proportional to the cosine of the angle between the geocentric location vector of the point of calculation and that of the dipole vector. Both values vanish because it is assumed that

- The total mass of the Earth $GM_\oplus$ assumed by the normal field is realistic.
- The origin of the co-ordinate reference system is assumed to be at the centre of mass of the Earth.

See section 3.4 for more.

Self-test questions

1. What is the basic idea behind using a normal gravity field?
2. What is the difference between gravity and gravitation?
3. Given the centrifugal potential

$$\Phi = \tfrac{1}{2}\omega_\oplus^2 \left(X^2 + Y^2\right),$$

derive the centrifugal acceleration as a vector. (X, Y, Z) are rectangular co-ordinates of a frame rotating at angular rate $\omega_\oplus$ around the Z axis.

---

[7] Earlier on we also used for this reference radius (in spherical approximation) the symbol R.

4. In figure 4.1 are drawn the equipotential surfaces of the normal gravity field. It is seen that they are farther apart over the equator than over the poles, because normal gravity over the equator is less than over the poles.

   How would the situation be for the normal *gravitational* field, that is, without the centrifugal force? Please explain your reasoning.

5. Explain the idea of natural co-ordinates.
6. What was the relationhip between M. Le Blanc and C. F. Gauss? Use Google.
7. Derive the Somigliana–Pizzetti equation 4.8 from equation 4.7. What makes the equation valuable?
8. What are the defining parameters of the Geodetic Reference System 1980?
9. Why does the spherical-harmonic expansion of the normal potential contain only a small number of terms and coefficients?
10. Why does the spherical-harmonic expansion of the normal potential not contain any terms of order $m \neq 0$?

järjestysluku

11. Why does the spherical-harmonic expansion of the normal potential contain only terms with even degree numbers n?

Exercise 4−1: The Somigliana–Pizzetti equation

1. Given gravity on the equator γ_a and on the poles γ_b, what is the gravity on geodetic latitude $\varphi = 45°$? Derive an expression that may also contain a and b.
2. And what is the gravity on the reduced latitude $\beta = 45°$? Compare with the previous.
3. Given are the semimajor axis a and semiminor axis b. What are the differences, for the same point, between the different latitude types (geodetic φ, geocentric ϕ, and reduced β) at most, in minutes of arc? Assume that the maximum happens at latitudes $\pm 45°$.

4. Compute for both a geodetic and a reduced latitude of 45° numerical values of gravity for the case of the GRS80 reference ellipsoid. By how much do they differ?

Exercise 4–2: Centrifugal force

Given is the rotation rate of the Earth in radians per second: $\omega_\oplus = 7292\,115 \cdot 10^{-11}\,\mathrm{s}^{-1}$.

1. Compute (roughly) the *centrifugal force* caused by the Earth's rotation at Southern Finland ($\varphi = 60°$, $R = 6378\,\mathrm{km}$, spherical Earth). In what direction does the force point (sketch!)?
2. How much does the centrifugal force contribute to local gravity, in other words, by how much does it change gravity, both as an acceleration and as a percentage?
3. Compute from the $\omega_\oplus$ value given above, the rotation time of the Earth in hours and minutes. Why is it not precisely 24^h?

Anomalous quantities of the gravity field

5

5.1 Disturbing potential, geoid height, deflections of the plumb line

The first *anomalous quantity*, which we already discussed above, is the difference between the true gravity potential W and the normal gravity potential U, the *disturbing potential*: häiriö-potentiaali

$$T \stackrel{\text{def}}{=} W - U.$$

All other anomalous quantities are various functions of the disturbing potential, such as the geoid height N and the plumb-line deflections ξ and η. They are generally obtained by subtracting from each other luotiviivan poikkeama

- a natural quantity related to the Earth's real gravity field, and
- a corresponding quantity related to the normal gravity field of the reference ellipsoid of the Earth. vertaus-ellipsoidi

For example, *deflections of the plumb line*

$$\xi \stackrel{\text{def}}{=} \Phi - \varphi, \quad \eta \stackrel{\text{def}}{=} (\Lambda - \lambda)\cos\varphi.$$

Here, (Φ, Λ) are astronomical latitude and longitude, that together make up the direction of the local plumb line, and (φ, λ) are the geodetic latitude and longitude that similarly make up the direction of the normal gravity vector or "normal plumb line".[1] See figure 5.1. luotiviiva [1]

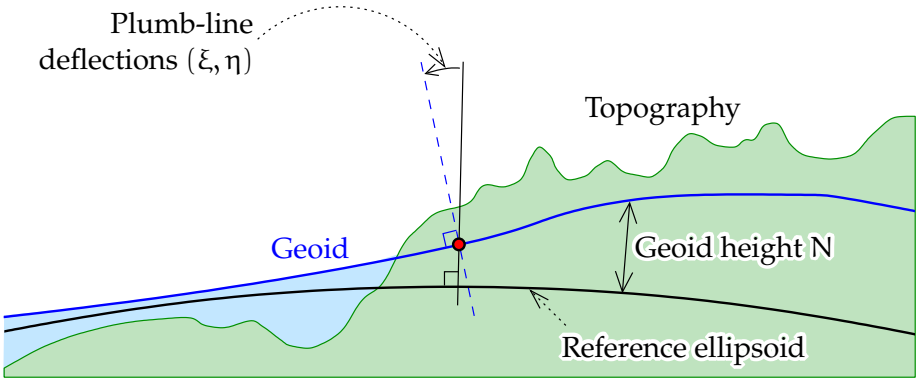


FIGURE 5.1. Geoid undulations N and deflections of the plumb line ξ and η.

The geoid height or *geoid undulation* is

$$N \stackrel{\text{def}}{=} h - H,$$

in which H is the orthometric height — reckoned from mean sea level — and h the height above the reference ellipsoid.

Deflections of the plumb line in Finland are a few seconds of arc (″) in magnitude. Geoid undulations range from 15 to 32 m (for comparison: globally the range of variation is −107 m to +85 m), relative to the GRS80 ellipsoid as is today customary. At sea level, the plumb-line deflections — expressed in radians! — equal the horizontal gradients of the geoid undulation. See figures 5.1 and 5.2.

For any reference ellipsoid, for example the GRS80 ellipsoid, there exists its own mathematically exact standard or *normal gravity field*, of which one equipotential or level surface is precisely that reference ellipsoid. Using this field, one may calculate for each gravity field quantity the corresponding normal quantity. By subtracting the normal quantity from the original one, the corresponding *anomalous* quantity is obtained.

[1] This assumes that Φ and Λ have been reduced to sea level, because of the curvature of the plumb line, section 4.3.2, and that φ and λ have been computed on the reference ellipsoid.

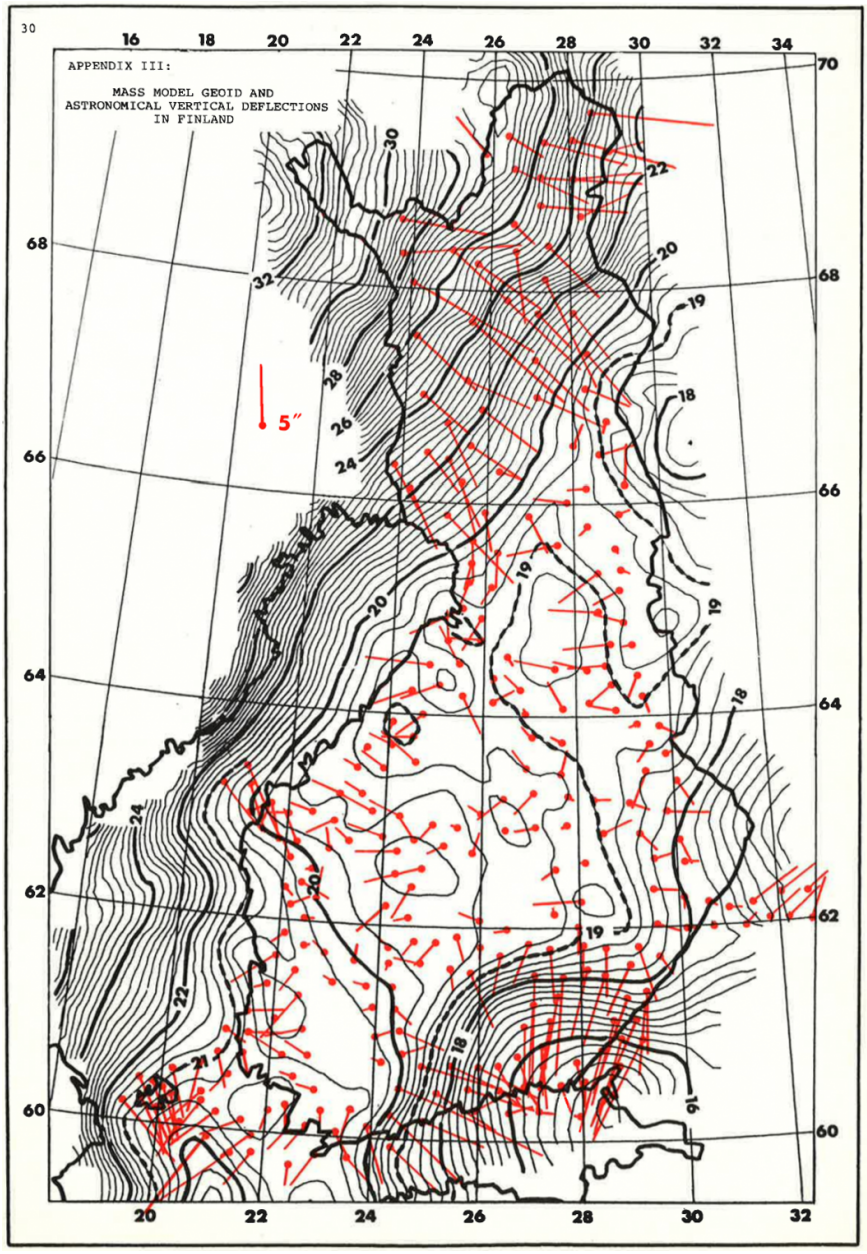


FIGURE 5.2. A geoid model for Finland from 1984. Deflections of the plumb line computed from observations in red (Vermeer, 1984).

For heights above the reference ellipsoid there exists an expression analogous to expression 4.5 for orthometric heights. Let U be the normal

potential and γ normal gravity:[2]

$$dU = -\gamma\, dh \quad \Longrightarrow \quad h_P = -\int_{U_0}^{U_P} \frac{1}{\gamma(U)} dU.$$

The geoid height in point P is now

$$\begin{aligned} N_P = h_P - H_P &= \int_{W_0}^{W_P} \frac{1}{g(W)} dW - \int_{U_0}^{U_P} \frac{1}{\gamma(U)} dU = \\ &= \int_{W_0}^{W_P} \frac{1}{g(W)} dW - \int_{W_0}^{W_P} \frac{1}{\gamma(U)} dU - \int_{W_P}^{U_P} \frac{1}{\gamma(U)} dU + \int_{W_0}^{U_0} \frac{1}{\gamma(U)} dU = \\ &= \int_{W_0}^{W_P} \frac{\gamma(W) - g(W)}{g(W)\,\gamma(W)} dW - \int_{W_P}^{U_P} \frac{1}{\gamma(U)} dU + \int_{W_0}^{U_0} \frac{1}{\gamma(U)} dU = \\ &= \int_0^{H_P} \frac{g(z) - \gamma(z)}{\gamma(z)} dz - \int_{W_P}^{U_P} \frac{1}{\gamma(U)} dU + \int_{W_0}^{U_0} \frac{1}{\gamma(U)} dU, \end{aligned} \tag{5.1}$$

by re-naming the integration variable $U \to W$ and changing it to a length: $dW = -g\, dz$.

In equation 5.1 the last term vanishes if we assume[3] $U_0 = W_0$. The first and second terms both depend on the height of point P, but their difference N_P does not. Therefore, instead of point P, we use its projection P′ to mean sea level — the zero point of the height system used. Then, the first term also vanishes: $H_{P'} = 0$. So

$$N_{P'} = -\int_{W_{P'}}^{U_{P'}} \frac{1}{\gamma(U)} dU \approx \frac{1}{\gamma_{P'}} (W_{P'} - U_{P'}) = \frac{T_{P'}}{\gamma_{P'}},$$

where we have substituted $T = W - U$, the disturbing potential. All quantities are now at sea level. More compactly:

$$N = \frac{T}{\gamma}. \tag{5.2}$$

This is the famous *Bruns*[4] equation (Heiskanen and Moritz, 1967, equa-

[2] This is not exactly true, due to the "normal plumb line" not being the same as the normal on the reference ellipsoid. The error made is tiny.

[3] This is not self-evident! In a local vertical datum the potential of the zero point could well differ by as much as the equivalent of a metre from the normal potential of a global reference ellipsoid.

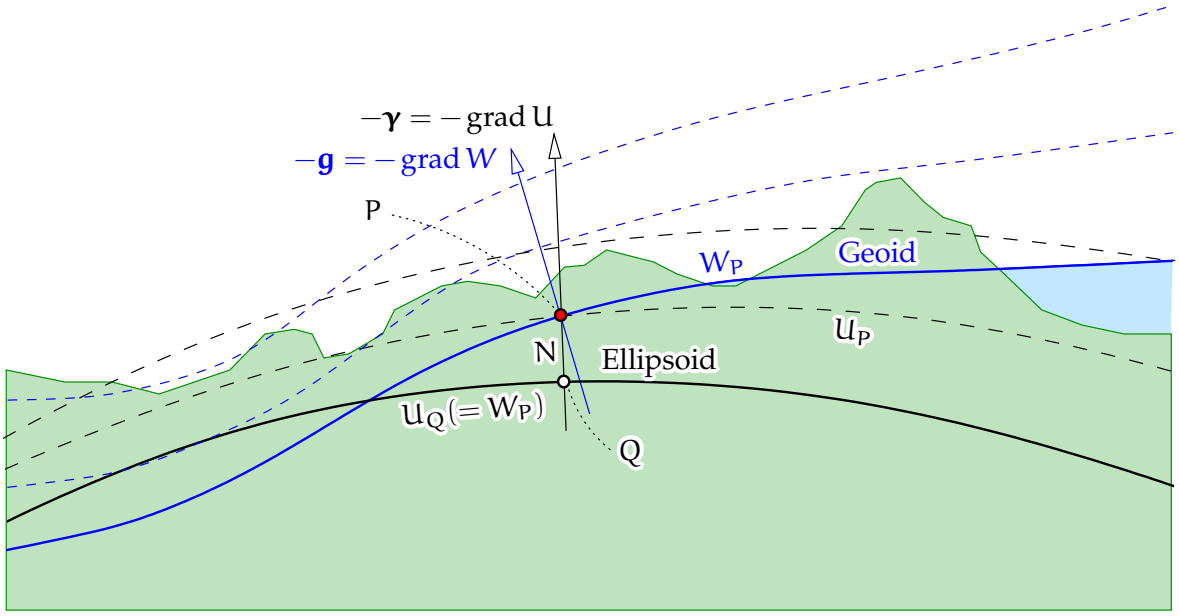


FIGURE 5.3. Equipotential or level surfaces of the gravity field (W) and the normal gravity field (U).

tion 2-144).

Figure 5.3 depicts the situation still better. In this figure, the normal gravity vector $\boldsymbol{\gamma} = \operatorname{grad} U$ has a length of $\gamma = \|\boldsymbol{\gamma}\| = -\frac{\partial}{\partial h} U$, from which it follows, with equation $T = W - U$, that the separation between "matching" surfaces $W = W_P$ and $U = U_Q$, when $W_P = U_Q$, is

$$N \approx \frac{U_Q - U_P}{\gamma} = \frac{W_P - U_P}{\gamma} = \frac{T}{\gamma}.$$

5.2 Gravity disturbances

The difference between the true and normal gravity accelerations is called the *gravity disturbance*, painovoima-häiriö

$$\delta g \overset{\text{def}}{=} g - \gamma = \|\mathbf{g}\| - \|\boldsymbol{\gamma}\| \approx -\left(\frac{\partial W}{\partial H} - \frac{\partial U}{\partial h}\right),$$

[4] Ernst Heinrich Bruns (1848–1919) was a German mathematician and mathematical geodesist.

in which the differentiation is done along the plumb line for W and — slightly imprecisely — along the normal on the reference ellipsoid for U. The directions of the plumb line and surface normal on the ellipsoid are actually very close to each other. Therefore, to good approximation

$$\delta g \approx -\left(\frac{\partial W}{\partial H} - \frac{\partial U}{\partial H}\right) = -\frac{\partial T}{\partial H}.$$

In spherical approximation we have

$$\delta g \approx -\frac{\partial T}{\partial r}. \tag{5.3}$$

asteosuus We already expanded the disturbing potential T into constituents for different spherical-harmonic degree numbers, equation 4.12, and now we obtain by differentiating with respect to r:

$$\delta g(\phi,\lambda,r) = -\frac{\partial T(\phi,\lambda,r)}{\partial r} = -\frac{\partial}{\partial r}\left(\sum_{n=2}^{\infty}\left(\frac{R}{r}\right)^{n+1} T_n(\phi,\lambda)\right) =$$

$$= \sum_{n=2}^{\infty}\frac{n+1}{r}\left(\frac{R}{r}\right)^{n+1} T_n(\phi,\lambda) = \sum_{n=2}^{\infty}\frac{n+1}{R}\left(\frac{R}{r}\right)^{n+2} T_n(\phi,\lambda), \tag{5.4}$$

and at sea level ($r = R$):

$$\delta g(\phi,\lambda,R) = \sum_{n=2}^{\infty}\frac{n+1}{R} T_n(\phi,\lambda).$$

This is the *spectral representation* of the gravity disturbance at sea level, on an Earth sphere of radius R. As a suitable value for the reference radius R one may take the equatorial radius $a_\oplus$ of a reference ellipsoid for the Earth.

We can determine gravity disturbances from observations only if, in addition to measuring the acceleration of gravity g_P $\left(= -\frac{\partial}{\partial H} W\big|_P\right)$ at a point P, we have a way to measure P's *location* in space, relative to the geocentre, so one may calculate normal gravity $\gamma_P = -\frac{\partial}{\partial h} U\big|_P$ at the same point. Nowadays this is even easy using GNSS, but traditionally it has been impossible. For this reason, gravity disturbances are little-used. One rather uses gravity *anomalies*, about which more in the following section.

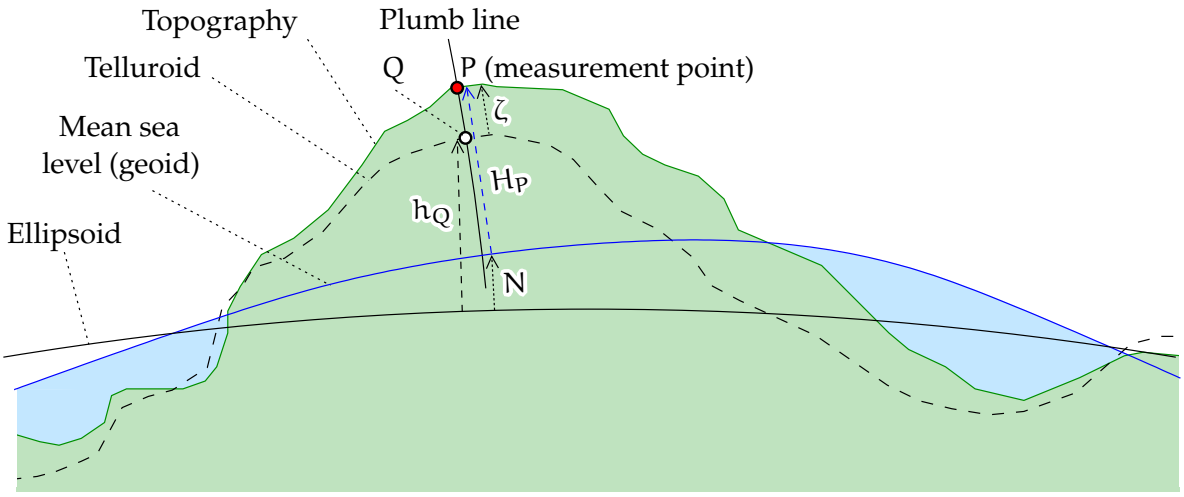


FIGURE 5.4. Reference ellipsoid, mean sea level (geoid), telluroid, and gravity measurement.

5.3 Gravity anomalies

Normal gravity is calculated as a function of location expressed in *geodetic* co-ordinates (φ, λ, h). However, in traditional gravimetric field work, before the satellite positioning era, one only had access to the geodetic co-ordinates φ and λ, but not to any accurate height h above the reference ellipsoid. One only had access to the height H above sea level (the geoid), determined, for example, through a national levelling network — or, in the worst case, barometrically.

This means that, although the true gravity g is measured at point P, the height of which above sea level is H_P, the normal gravity γ must of necessity be calculated *at another point* Q, the height of which *above the reference ellipsoid* is $h_Q = H_P$. See figure 5.4.

In other words, the measured height of point P *above mean sea level* is substituted, brute-force style, into the normal gravity formula, which, however, expects a height *above the reference ellipsoid*! This special trait of the definition of gravity anomalies may be called a "free boundary-value problem". reuna-arvo-tehtävä

According to this, we derive an expression for the gravity anomaly as follows:

$$\begin{aligned}\Delta g_P = g_P - \gamma_Q = (g_P - \gamma_P) + (\gamma_P - \gamma_Q) = \\ = -\left(\frac{\partial W}{\partial H}\Big|_P - \frac{\partial U}{\partial h}\Big|_P\right) + (\gamma_P - \gamma_Q) \approx \\ \approx -\frac{\partial (W-U)}{\partial H}\Big|_P + (h_P - h_Q)\,\frac{\partial \gamma}{\partial H}\Big|_P = \\ = -\frac{\partial T}{\partial H}\Big|_P + (h_P - H_P)\,\frac{\partial \gamma}{\partial H}\Big|_P = \\ = -\frac{\partial T}{\partial H}\Big|_P + N_P\,\frac{\partial \gamma}{\partial H}\Big|_P = \left(-\frac{\partial T}{\partial H} + \frac{T}{\gamma}\frac{\partial \gamma}{\partial H}\right)\Big|_P,\end{aligned}$$

using almost all the equations above. This equation,

$$\Delta g = -\frac{\partial T}{\partial H} + \frac{1}{\gamma}\frac{\partial \gamma}{\partial H}T, \tag{5.5}$$

fysikaalisen geodesian perusyhtälö

is known as the *fundamental equation of physical geodesy*. It is the boundary condition of the *third boundary-value problem* (Heiskanen and Moritz, 1967, section 1-17). It enables the solution of T in the exterior space, if Δg is given everywhere on the Earth's surface.

If we assume that the Earth is a sphere of radius R and that the normal gravity field is spherically symmetric, we may approximate:

$$\Delta g = -\frac{\partial T}{\partial r} - \frac{2}{r}T, \tag{5.6}$$

in which $r = R + H$ is the distance from the Earth's centre.

By substituting into this equation 5.3 for δg we obtain

$$\Delta g = \delta g - \frac{2}{r}T.$$

From this we obtain by substituting the spectral representations 3.10

(but for T) and 5.4 for δg:

$$\Delta g(\phi,\lambda,r) = \sum_{n=2}^{\infty}\left(\frac{n+1}{r}-\frac{2}{r}\right)\left(\frac{R}{r}\right)^{n+1}T_n(\phi,\lambda) =$$
$$= \sum_{n=2}^{\infty}\frac{n-1}{r}\left(\frac{R}{r}\right)^{n+1}T_n(\phi,\lambda) = \sum_{n=2}^{\infty}\frac{n-1}{R}\left(\frac{R}{r}\right)^{n+2}T_n(\phi,\lambda) =$$
$$= \sum_{n=2}^{\infty}\left(\frac{R}{r}\right)^{n+2}\Delta g_n(\phi,\lambda), \quad (5.7)$$

with the notation

$$\Delta g_n(\phi,\lambda) \stackrel{\text{def}}{=} \frac{n-1}{R}T_n(\phi,\lambda). \quad (5.8)$$

The presence of the factor $n-1$ shows that gravity anomalies cannot contain constituents of degree $n=1$, even if T would contain them. It is always wise to place the origin of the co-ordinate system in the centre of mass of the Earth, but if the origin is not located there, at least gravity anomalies do not change.

At sea level $r = R$ we find

$$\Delta g(\phi,\lambda,R) = \sum_{n=2}^{\infty}\Delta g_n(\phi,\lambda),$$

showing the Δg_n to be the *degree constituents* of gravity anomaly Δg.

Observe that the term $n=1$ is missing: $\Delta g_1 = 0$. It is also assumed that $\Delta g_0 = -T_0/R = 0$, meaning that the true exterior potential is on global average the same as the normal potential. Also the total mass $GM_\oplus$ and geoid volume[5] of the Earth are assumed to be the same as the total mass and volume of the reference ellipsoid. The assumption is largely justified because $GM_\oplus$ can be, and has been, determined very precisely by satellites, and modern models for the normal potential, like GRS80, are based on these determinations.[6]

[5] In fact, the atmosphere complicates this matter.

[6] However, GRS80 has an equatorial radius of 6 378 137.0 m, while the newer models

TABLE 5.1. Orders of magnitude of gravity variations.

Phenomenon	Fraction of gravity	SI units	mGal
Ambient gravity	1	9.81	981 000
Variation with location	$\pm 10^{-4}$	$\pm 10^{-3}$	± 100
Difference equator – poles	0.5 %	0.05	5000
Difference sea surface – 10 km high	0.3 %	0.03	3000
Gravimeter accuracy	$\pm\, 10^{-8}$–10^{-7}	$\pm\, 10^{-7}$–10^{-6}	$\pm\, 0.01$–0.1

5.4 Units used for gravity anomalies

A common unit of measurement for gravity variations is the *milligal*. The connection with the SI system is 1 mGal $= 10^{-5}$ $\mathrm{m/s^2}$. The unit µGal or 10^{-8} $\mathrm{m/s^2}$ is also used. The units $\mathrm{\mu m/s^2}$ and $\mathrm{nm/s^2}$, which formally belong to the SI system, are also used in modern books. Nevertheless, milligals and microgals are more familiar still, and correspond to 1 ppm (part per million) and 1 ppb (part per billion) of ambient gravity close to the Earth's surface.

In table 5.1 we give a few values in order to get an idea of the orders of magnitude of phenomena.

A popular unit for measuring gravity *gradients* is the eötvös, symbol E. In SI units it is 10^{-9} $\mathrm{s^{-2}}$, corresponding to 10^{-4} $\mathrm{mGal/m}$. On the Earth's surface the vertical gradient $\frac{\partial}{\partial H} g$ is on average some -0.3 $\mathrm{mGal/m} = -3000$ E.

5.5 The boundary-value problem of physical geodesy

As explained in the above section, gravimetric measurement is more complicated than just measuring the quantity $-\frac{\partial}{\partial H} W \approx -\frac{\partial}{\partial r} W$. When we measure the vertical derivative of the geopotential, we do it *in a place*

like EGM2008 give a smaller value of 6 378 136.3 m as the mean location of global mean sea level. This is something to be well aware of when using the model in production work. Uncertainty continues to be of decimetre order.

we do not precisely know. Even if we know the height of the measurement location above sea level, that still does not give us the measurement point's location in space. This location depends additionally on the location of sea level, the *geoid*, in space; more precisely its height above or below the reference ellipsoid.

This is how we arrive at the third boundary-value problem.[7] The *boundary-value problem of physical geodesy* is to determine the potential V outside a body if given on its surface is the linear combination [7]

$$c_1 V + c_2 \frac{\partial V}{\partial n},$$

with c_1 and c_2 suitable coefficients. The variable n represents here differentiation in the direction of the normal to the boundary surface, in practice the same as H or r.

In physical geodesy, the following linear combination is given as a boundary condition:

$$\Delta g = -\frac{\partial T}{\partial H} + \frac{1}{\gamma}\frac{\partial \gamma}{\partial H} T. \tag{5.5}$$

We see that $c_1 = -1$ and $c_2 = \gamma^{-1}\frac{\partial}{\partial H}\gamma$. This equation, the definition 5.5 of gravity anomalies, is known as the *fundamental equation of physical geodesy*.

Again in spherical approximation, inverting equation 5.8 yields

$$T_n(\phi, \lambda) = \frac{R}{n-1}\Delta g_n(\phi, \lambda).$$

Remember that the functions $\Delta g_n(\phi, \lambda)$ are computable using the degree constituent equation 3.9 when $\Delta g(\phi, \lambda)$ is known all over the Earth. asteosuus-yhtälö

This is how we also obtain the solution of this boundary-value problem in spectral representation — which is thus valid in the *whole*

[7] The third or mixed boundary-value problem is associated with Victor Gustave Robin (1855–1897), a French mathematician. Then, the Dirichlet problem could be called the first and the Neumann problem the second boundary-value problem.

exterior space:

$$T(\phi,\lambda,r) = \sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+1} T_n(\phi,\lambda) = \sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+1} \frac{R}{n-1} \Delta g_n(\phi,\lambda) =$$
$$= \frac{R}{4\pi} \sum_{n=2}^{\infty} \frac{2n+1}{n-1} \left(\frac{R}{r}\right)^{n+1} \iint_{\sigma} \Delta g(\phi',\lambda',R)\, P_n(\cos\psi)\, d\sigma'. \quad (5.9)$$

This is precisely the boundary-value problem that is created if surface gravity anomalies are given everywhere on the Earth, land and sea.

The integral equation corresponding to the above spectral equation 5.9 is known as the Stokes[8] equation:

$$T(\phi,\lambda,r) = \frac{R}{4\pi} \iint_{\sigma} S(\psi,r,R)\, \Delta g(\phi',\lambda',R)\, d\sigma',$$

in which the *Stokes kernel* is

$$S(\psi,r,R) = \sum_{n=2}^{\infty} \frac{2n+1}{n-1} \left(\frac{R}{r}\right)^{n+1} P_n(\cos\psi). \quad (5.10)$$

In section 8.1 we will give closed form 8.3 for this function for the case $r = R$, and a graph.

5.6 The telluroid mapping and the "quasi-geoid"

If we measure the astronomical latitude and longitude (Φ,Λ) and *interpret* them as geodetic (geographical) co-ordinates (φ,λ), and also *interpret* the potential difference $-(W-W_0)$ as a measure for the height above the reference ellipsoid h, we perform, as it were, a *mapping*. This mapping adds to every point P a corresponding point Q, the *geodetic* co-ordinates of which are the same as the *natural* co-ordinates of point P.

This approach is called *telluroid mapping*. The telluroid is the surface that follows the shapes of the Earth's topography, but is everywhere

[8] Sir George Gabriel Stokes PRS (1819–1903) was an Irish-born, gifted mathematician and physicist who made his career in Cambridge.

below the topography by an amount ζ if positive, or above it by an amount $-\zeta$ otherwise. The quantity ζ is called a *height anomaly*.

Telluroid mapping is an important tool in Molodensky's gravity field theory. It is, however, a pretty abstract concept. One may say that the telluroid is a *model* of the Earth's surface, obtained by assuming that

- The true potential field of the Earth is the normal potential.
- The mathematical mean sea surface or *geoid*, the reference surface for height measurement, coincides with the reference ellipsoid.

In other words, the telluroid is a model for the Earth's topographic surface that is obtained by taking *levelled* heights — more precisely, geopotential numbers obtained from levelling — as if they represented differences of the *normal* potential from that of the reference ellipsoid.

In practice, a map of values ζ is often called a "quasi-geoid" model. The quasi-geoid is usually close to the geoid, except in the mountains, where the differences can exceed a metre.

One should however remember that the height anomaly ζ is defined *on the topographic surface*; a surface that is quite rough in many places. This means that all variations in topographic height will also be reflected as variations in this quasi-geoid, in such a way that the quasi-geoid *correlates* strongly with the small details in the topography. One can thus not say that the shape of the quasi-geoid only expresses the figure of the Earth's potential field. In it, variations in geopotential and in topographic height are hopelessly mixed up.

This is why the quasi-geoid is an unfortunate compromise, a concession to "reference-surface thinking", which only really works within the classical geoid concept. Better stick — within Molodensky's theory — to the concept of *height anomaly*, which is a three-dimensional function or field

vertauspinta

$$\zeta(X, Y, Z) = \zeta(\varphi, \lambda, h).$$

5.7 Free-air anomalies

If we measure gravity g at point P, the height of which over sea level is H and the latitude Φ, we may calculate the gravity anomaly Δg at the point as follows:

$$\Delta g \stackrel{\text{def}}{=} g - \gamma(\Phi, H),$$

in which $\gamma(\Phi, H)$ is normal gravity calculated according to its formal definition, but at height H and latitude Φ.

ilma-anomalia This is how we define free-air anomalies.

We *linearise* this as follows:

$$\Delta g = g - \gamma(\Phi, H) \approx g - \left(\gamma(\varphi, h) + (\Phi - \varphi)\,\frac{\partial \gamma}{\partial \varphi} + (H - h)\,\frac{\partial \gamma}{\partial h}\right) \approx$$
$$\approx g - \left(\gamma(\varphi, 0) + h\frac{\partial \gamma}{\partial h} + (H - h)\,\frac{\partial \gamma}{\partial h}\right) = g - \gamma(\varphi, 0) - H\frac{\partial \gamma}{\partial h},$$

making the approximation that the vertical gradient $\frac{\partial}{\partial h}\gamma$ of normal gravity is constant.[9]

Thus, free-air anomalies can be calculated in a simpler way. The gravity formula of the normal field 4.10 gives for latitude 60°:

$$\gamma = 981\,917.838 - 0.308\,449\,4\,H + \cdots \text{ mGal}.$$

So, in linear approximation (close to the Earth's surface) gravity attenuates some 0.3 mGal for every metre in height. This value is worth committing to memory.

An approximate equation for calculating free-air anomalies then is

$$\Delta g_P = g_P - \gamma_0(\varphi) + 0.3084\,{}^{\text{mGal}}/_{\text{m}}\,H, \tag{5.11}$$

[9] For the greatest precision, one should consider that the latitude Φ may also not be a latitude on a geocentric reference ellipsoid. It could be astronomical latitude, or latitude in some old national co-ordinate system computed on a non-geocentric ellipsoid, like in Finland KKJ, the National Map Grid Co-ordinate System, which was computed on the Hayford ellipsoid. The error caused by this is however two, three orders of magnitude smaller than the effect caused by $H - h$.

in which $\gamma_0(\varphi) \stackrel{\text{def}}{=} \gamma(\varphi, 0)$, normal gravity at sea level, is only a function of latitude. In a country like Finland, equation 5.11 is often sufficiently precise, although the evaluation of the original equation 4.10 is also easy.

Free-air anomalies are widely used. Generally, when one discusses gravity anomalies, one means just this, free-air anomalies. They express the Earth's exterior gravity field, including mountains, valleys and everything.

Questions

1. If gravity at sea level is $9.81\,{}^{\mathrm{m}}/_{\mathrm{s}^2}$, at what height will gravity disappear, as computed according to the above-mentioned vertical gravity gradient $-0.3\,{}^{\mathrm{mGal}}/_{\mathrm{m}}$?
2. How physically realistic is this?

Answers

1. At $-0.3\,{}^{\mathrm{mGal}}/_{\mathrm{m}}$, it takes $\left(9.81 \cdot 10^5 \,/\, 0.3\right)$ m $= 3270$ km to go to zero.
2. Not very. The gravity gradient itself drops quickly from the value of $-0.3\,{}^{\mathrm{mGal}}/_{\mathrm{m}}$ going up, so this linear extrapolation is simply wrong.

Self-test questions

1. How do deflections of the plumb line and geoid heights relate to each other?
2. What is the fundamental equation of physical geodesy in spherical approximation?
3. In what way is a gravity disturbance different from a gravity anomaly?
4. What units are used for measuring gravity anomalies and gravity

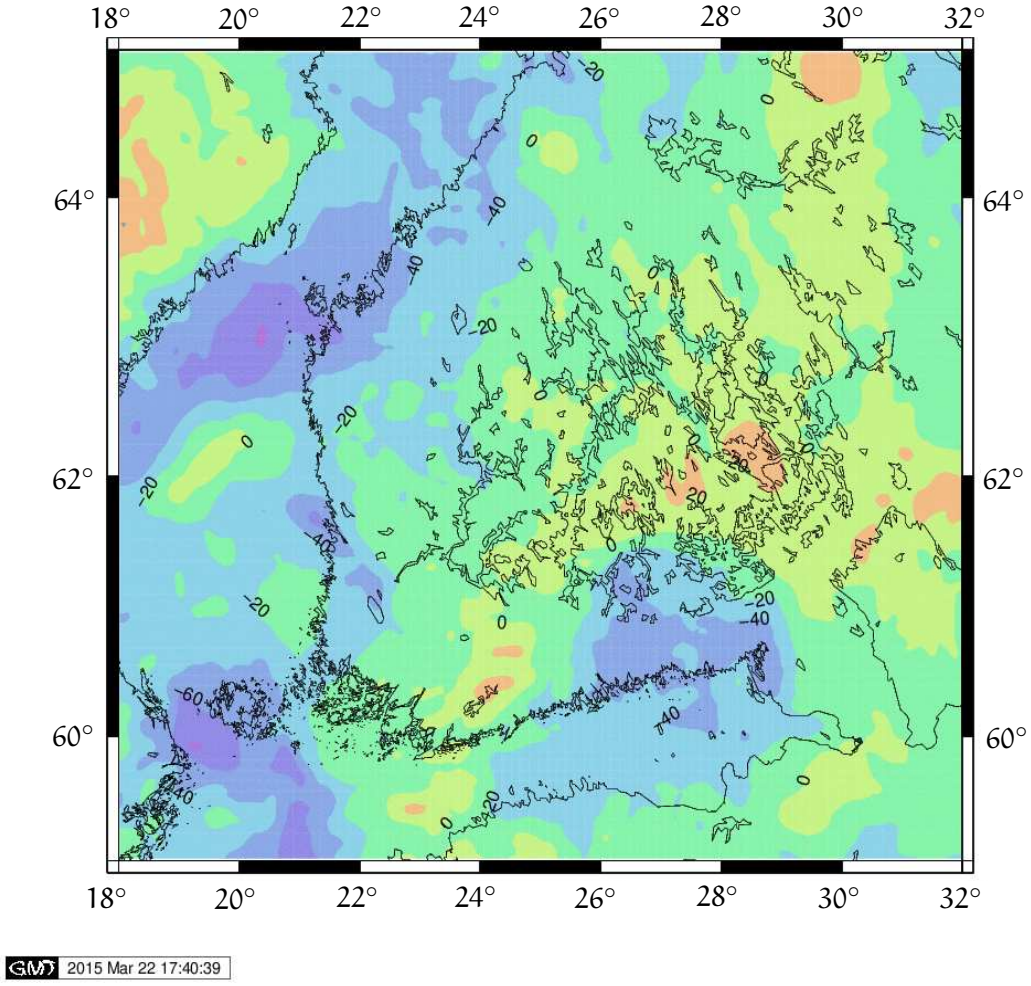


FIGURE 5.5. Free-air gravity anomalies for Southern Finland, computed from the EGM2008 spherical-harmonic expansion. Data © Bureau Gravimétrique International (BGI) / International Association of Geodesy. Web service BGI, EGM2008.

gradients? How are they related to the SI system?

5. How do the geoid height and the disturbing potential relate to each other?
6. Explain telluroid mapping and height anomalies.

Exercise 5−1: The spectrum of gravity anomalies

Use equation 5.8. If we assume that the quadratic mean of the degree constituents Δg_n of gravity anomalies,

$$\|\Delta g_n\|_{\overline{\sigma}} \stackrel{\text{def}}{=} \sqrt{\frac{1}{4\pi} \iint_{\sigma} \Delta g_n^2 \,(\phi, \lambda)\, d\sigma},$$

does not depend on the chosen degree number n, how then does the similarly defined $\|T_n\|_{\overline{\sigma}}$ depend on n?

In other words: which degree numbers of the gravity field are relatively strongest in the disturbing potential, and which in the gravity anomalies?

Exercise 5−2: Deflections of the plumb line and geoid tilt

If, in the south-north components of plumb-line deflections in a country, there is a systematic error of one arc second, what error does this cause in the difference $N_2 - N_1$ between the geoid heights in points 1 and 2 of which the inter-point distance is approximately 1000 km in the south-north direction? See figures 5.1 and 5.2.

Exercise 5−3: Gravity anomaly, geoid height

In Finland there is a place where the (free-air) gravity anomaly is $\Delta g = 100\,\text{mGal} = 10^{-3}\,\text{m/s}^2$. In the same place the disturbing potential T is $200\,\text{m}^2/\text{s}^2$.

1. Using the fundamental equation of physical geodesy 5.6:

$$\Delta g = -\frac{\partial T}{\partial r} - \frac{2}{r} T,$$

calculate $\frac{\partial}{\partial r} T$ and compare it with the quantity $2T/r$. Assume $r \approx R$. Which of the two, $\frac{\partial}{\partial r} T$ or $2T/r$, dominates?

2. Assume that the point is close to sea level. Using the Bruns equation

$$N = \frac{T}{\gamma},$$

where γ is average gravity $9.81\,{}^{m}/_{s^2}$, compute the geoid height N of the point.

Geophysical reductions

6

6.1 General

We see that integral equations, like Green's third theorem 1.25, offer a possibility to calculate the whole exterior potential of the Earth — as well as all the quantities that may be calculated from the potential, such as, for example, the gravitational acceleration — using values of V or $\frac{\partial}{\partial n}V$ — or a linear combination — observed on the boundary surface only. A requirement is, that there are no masses outside the boundary surface.

Green's third theorem is but one example of many: every integral theorem is the solution of some *boundary-value problem*.

reuna-arvo-tehtävä

There are three alternatives for choosing a boundary surface:

1. Choose the topographic surface of the Earth.
2. Choose mean sea level, more precisely, an equipotential surface close to mean sea level called the *geoid*.
3. Choose the reference ellipsoid.

vertaus-ellipsoidi

   - Alternative 1 has been developed mostly by the Molodensky school (Molodensky et al., 1962) in the former Soviet Union. The advantage of the method is that *we need no gravity reduction*, as all masses are already inside the boundary surface. Its disadvantage is that the, often complex, shape of the topography must be taken

into account when the boundary-value problem is formulated and solved.

- Alternative 2 is classical geoid or geopotential determination. In this case *geophysical reductions* are needed to the input gravity data: some masses are outside the computation boundary and need to be computationally removed or moved to the inside. Only then will the Laplace equation 1.13 apply in the exterior space of the Earth, as required by the boundary-value problem of physical geodesy, see section 5.5.

  Then, the geopotential or geoid solution obtained is not any more that of the original mass distribution, but of the *reduced* one. This surface is called the *co-geoid*. We need a "restoration step" where this influence of the reduction step on the geopotential and geoid is determined and reversed. This influence is called the "indirect effect".

  epäsuora vaikutus

  poistamis-entistämis-menetelmä

  In the literature, this method is also referred to as a *remove–restore* method.

- Alternative 3 has been used rarely, because it has not been traditionally possible to make gravity measurements in a location known in the absolute sense, relative to the geocentre or the reference ellipsoid. Nowadays this is possible using GNSS, for example in Antarctica and Greenland's interior, where there is no sea-level-bound height system.

  We may expect this approach to gain more traction as heights of gravimetric stations are more and more determined directly using GNSS. See for example Märdla (2017).

6.2 Bouguer anomalies

ilma-anomalia

Free-air anomalies depend on the topography, because gravity itself contains the attractive effect of topographic masses. A map of free-air anomalies shows the same small details as seen in the topography. One

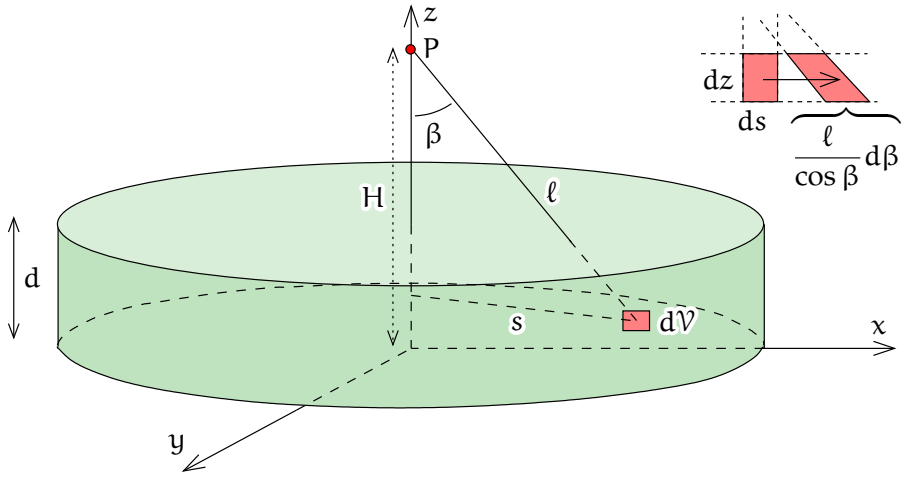


FIGURE 6.1. The attraction of a Bouguer plate.

way of removing the effect of the topography is the so-called *Bouguer*[1] *reduction*. [1]

6.2.1 Calculation of the Bouguer reduction

We calculate the effect of a homogeneous plate on gravity. Assume that the plate is infinite in size; thickness d, matter density ρ, and height H of point P above the lower surface of the plate. See figure 6.1. The attraction at point P, which is directed straight downwards for reasons of symmetry, is obtained by integrating. The volume integral to be computed has a volume element

$$dV = ds \cdot dz \cdot s\, d\alpha$$

in the cylindrical co-ordinates (s, z, α). *Transform* this to the co-ordinates (β, z, α). Forget about α and study the surface element (figure 6.1, top

[1] Pierre Bouguer (1698–1758) was a French professor in hydrography, who participated in the public discussion on the figure of the Earth, and in 1735–1743 led an expedition of the French Academy of Sciences doing a grade measurement in Peru, South America, at the same time as De Maupertuis was carrying out a similar grade measurement in Lapland. In addition to geodesy, he was also active in astronomy.

right)

$$\mathrm{ds}\,\mathrm{dz} = \frac{\ell}{\cos\beta}\mathrm{d}\beta\,\mathrm{dz},$$

in which the determinant of Jacobi needed, $\ell/\cos\beta$, is seen.

Carry out the integration:

$$\begin{aligned} a \overset{\text{def}}{=} \|\mathbf{a}\| &= G\iiint \frac{\cos\beta}{\ell^2}\rho\,d\mathcal{V} = G\rho\int_0^{2\pi}\int_0^{d}\int_0^{\infty}\frac{\cos\beta}{\ell^2}\cdot \mathrm{ds}\,\mathrm{dz}\cdot s\,d\alpha = \\ &= G\rho\int_0^{2\pi}\int_0^{d}\int_0^{\pi/2}\frac{\cos\beta}{\ell^2}\cdot\frac{\ell}{\cos\beta}\mathrm{d}\beta\,\mathrm{dz}\cdot s\,d\alpha = \\ &= 2\pi G\rho\int_0^{d}\int_0^{\pi/2}\frac{s}{\ell}\mathrm{d}\beta\,\mathrm{dz} = 2\pi G\rho\int_0^{d}\left(\int_0^{\pi/2}\sin\beta\,\mathrm{d}\beta\right)\mathrm{dz}. \end{aligned}$$

Here, the integral

$$\int_0^{\pi/2}\sin\beta\,\mathrm{d}\beta = \left[-\cos\beta\right]_0^{\pi/2} = 1,$$

and the end result is

$$a = 2\pi G\rho d. \tag{6.1}$$

This is the formula for the attraction of a Bouguer plate. As a side result, we obtain the attraction of a circular disk of radius r:

$$\int_0^{\beta_0(z)}\sin\beta\,\mathrm{d}\beta = \left[-\cos\beta\right]_0^{\beta_0(z)} = 1-\cos\bigl(\beta_0(z)\bigr),$$

and the whole integral

$$a = 2\pi G\rho\int_0^{d}\left(1-\overbrace{\frac{H-z}{\sqrt{(H-z)^2+r^2}}}^{\cos(\beta_0(z))}\right)\mathrm{dz}.$$

The indefinite integral is

$$\int\frac{H-z}{\sqrt{(H-z)^2+r^2}}\mathrm{dz} = -\sqrt{(H-z)^2+r^2}.$$

Substituting the bounds yields

$$\int_0^{d}\left(1-\frac{H-z}{\sqrt{(H-z)^2+r^2}}\right)\mathrm{dz} = d+\sqrt{(H-d)^2+r^2}-\sqrt{H^2+r^2}.$$

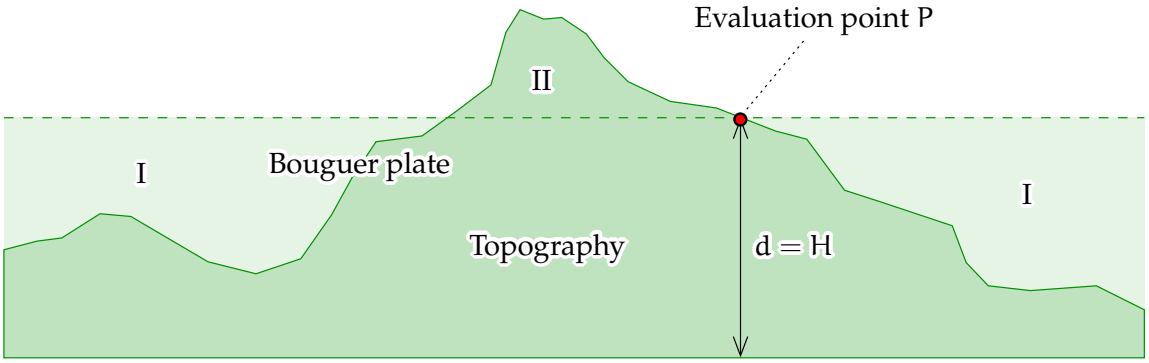


FIGURE 6.2. The Bouguer plate as an approximation to the topography.

We obtain for the whole integral

$$a = 2\pi G\rho \left(d + \sqrt{(H-d)^2 + r^2} - \sqrt{H^2 + r^2} \right).$$

In the limit $r \to \infty$, and thus

$$\sqrt{(H-d)^2 + r^2} - \sqrt{H^2 + r^2} \to 0,$$

this is identical to equation 6.1.

Bouguer anomalies are computed in order to remove the attraction of masses of the Earth's crust above sea level, above the *geoid*. The true topography is *approximated* by a Bouguer plate, see figure 6.2.

There is no standard way to treat sea-covered areas:

- Some maps show Bouguer anomalies over land and free-air anomalies over the sea. This is an option if no good quality depth data is available.
- A more correct way is to replace sea water by a rocky Bouguer plate, the thickness of which equals the local sea depth or *bathymetry*.

The calculation goes as follows:

$$\Delta g_B = \Delta g_{FA} - 2\pi G\rho H = \Delta g_{FA} - 0.1119\,H, \tag{6.2}$$

where we assume for the density ρ of the plate an often-used value for the average density of the Earth's crust, $\rho = 2670\,\mathrm{kg/m^3}$. By substituting

equation 5.11 into this, we obtain

$$\Delta g_B = g_P - \gamma_0(\varphi) + (0.3084 - 0.1119)H = g_P - \gamma_0(\varphi) + 0.1965\,H. \tag{6.3}$$

The quantity Δg_B is called a (simple) *Bouguer anomaly*.

The difference between the attraction of a Bouguer plate and that of the true topography is called the *terrain correction* TC (volumes I and II in figure 6.2). We shall return to its calculation later.

6.2.2 Properties

Unlike free-air anomalies which vary on both sides of zero, Bouguer anomalies are *strongly negative*, especially in the mountains. For example, if the mean elevation of a mountain range is $\overline{H} = 1000$ m, the Bouguer anomalies will, as a consequence of this, contain a *bias* of $1000 \times (-0.1119\,\text{mGal}) = -112\,\text{mGal}$, about -100 mGal for every kilometre of elevation.

systematiikka

The advantage of Bouguer anomalies is their smaller variation with place. For this reason they are suited especially for the *interpolation* and *prediction* of gravity anomalies, in situations where the available gravimetric material is geographically sparse. This requires that topographic heights are known to a better spatial density.

6.3 Terrain effect and terrain correction

Using the simple Bouguer reduction does not remove precisely from gravity anomalies the attractive effect of the whole topography. Figure 6.2 shows that we make two types of error:

- The attraction of volumes I is taken along, although there is nothing there.
- The attraction of volumes II, where there actually is something, is ignored.

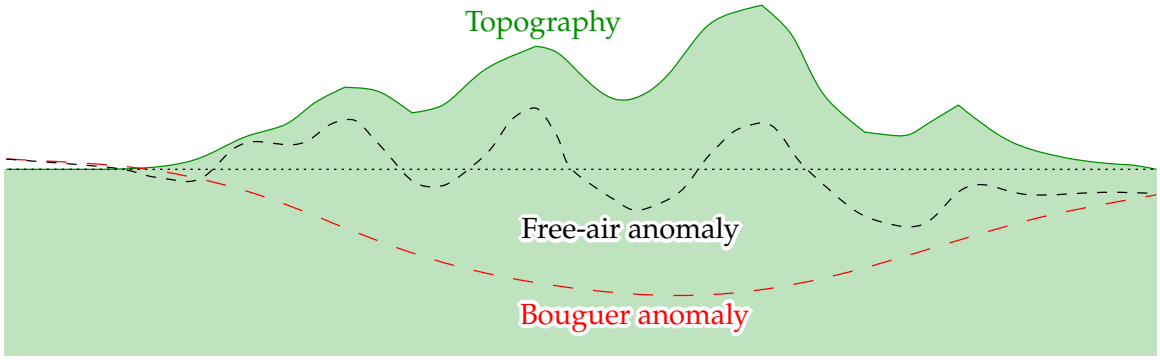


FIGURE 6.3. Behaviour of different anomaly types in mountainous terrain.

Both errors work in the same direction! Because volumes I are below the point of evaluation, their attraction — which the simple Bouguer reduction considers present, and removes — acts downwards. And because volumes II are above the point of evaluation, their attraction — which in the simple Bouguer reduction is not corrected for — acts upwards. The error made is in the same direction as in the previous case.

The terrain correction is always positive.

We write

$$\Delta g'_B = \Delta g_B + TC,$$

where TC — the "terrain correction" — is positive. $\Delta g'_B$ is called the *terrain-corrected Bouguer anomaly*.

The terrain correction is calculated by numerical integration. Figure 6.5 shows the *prism integration method*, and how both prisms, I and II, lead to a positive correction, because prism I is computationally added and prism II removed. One needs a digital terrain model, DTM, which must be, especially around the evaluation point, extremely dense: according to experience, 500 m is the maximum inter-point separation in a country like Finland, and in the mountains one needs even 50 m. The systematic nature of the terrain correction makes a too-sparse terrain model cause,

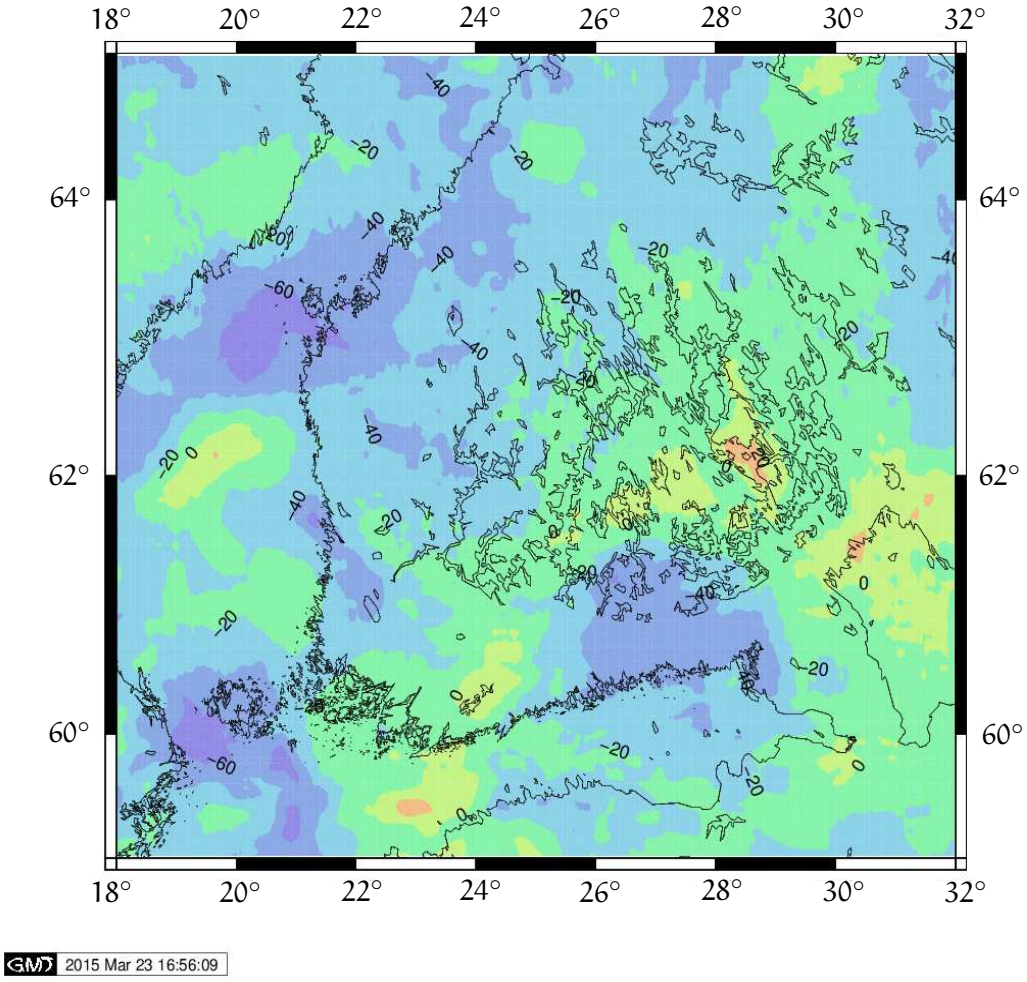


FIGURE 6.4. Terrain-corrected Bouguer anomalies for Southern Finland, computed from the spherical-harmonic expansion EGM2008. Data © Bureau Gravimétrique International (BGI) / International Association of Geodesy. Web service BGI, EGM2008. In comparison to figure 5.5 on page 126, there is a strong negative bias of Bouguer anomalies — although part of this is due to post-glacial isostatic imbalance and also visible in the free-air map. Bouguer anomalies are also smoother, but that is harder to see here, as Southern Finland is already a smooth area.

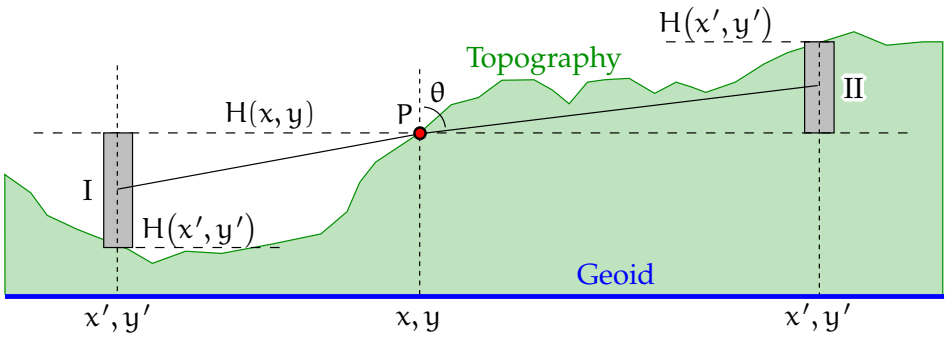


FIGURE 6.5. Calculating the classical terrain correction using the prism method.

possibly serious, *biases* in the insufficiently corrected gravity anomalies.

To compute the terrain correction with the prism method, we use the following equation, assuming a constant crustal density ρ and a flat Earth, in rectangular map co-ordinates x, y:

$$\mathrm{TC}(x,y) = \tfrac{1}{2} G\rho \int_{-D}^{+D}\int_{-D}^{+D} \frac{1}{\ell^3}\Big(H(x',y') - H(x,y)\Big)^2 \,dx'\,dy',$$

in which

$$\ell = \sqrt{(x-x')^2 + (y-y')^2 + \Big(\tfrac{1}{2}\big(H(x',y') - H(x,y)\big)\Big)^2}$$

is the distance between the evaluation point

$$\begin{bmatrix} x & y & H(x,y) \end{bmatrix}^{\mathsf{T}}$$

and the centre point on the central axis of the prism

$$\begin{bmatrix} x' & y' & \tfrac{1}{2}\Big(H(x',y') + H(x,y)\Big) \end{bmatrix}^{\mathsf{T}}.$$

Of course, this is only an approximation, but it works well enough in terrain where slopes generally do not exceed 45°. In the integral above, the limit D is typically tens or hundreds of kilometres. In the latter case, the curvature of the Earth already starts having an effect, which the equation does not consider.

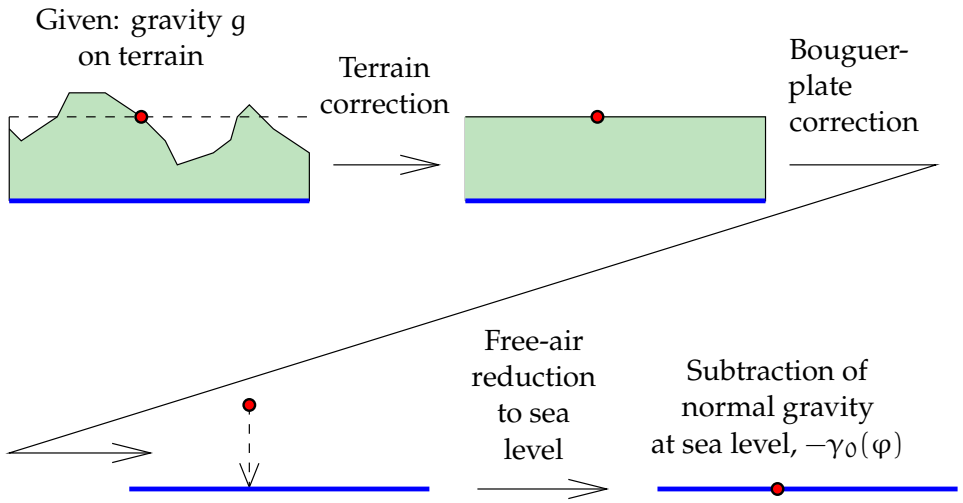


FIGURE 6.6. The steps in calculating the Bouguer anomaly. The reduction to sea level uses the standard free-air vertical gravity gradient, $-0.3084\,^{\mathrm{mGal}}/_{\mathrm{m}}$, the vertical gradient of normal gravity.

The values of the terrain correction TC vary from fractions of a milligal (Southern Finland) to hundreds of milligals (high mountain ranges). käsivarsi In the "arm" of Finland — the north-western, somewhat mountainous border area with Sweden and Norway — the terrain correction may be tens of milligals.

Figure 6.6 shows the stages of calculating Bouguer anomalies from gravity observations through terrain correction, Bouguer-plate correc- ilmareduktio tion and free-air reduction.

6.3.1 *Example: applying the terrain correction in a special case*

Given is the special terrain shape rendered in quasi-3-D in figure 6.7. The height differences are $PQ' = 300\,\mathrm{m}$ and $QQ' = 200\,\mathrm{m}$. The rock density is the standard crustal density, $2670\,^{\mathrm{kg}}/_{\mathrm{m}^3}$.

Questions

1. Calculate the terrain correction at point P (hint: use the

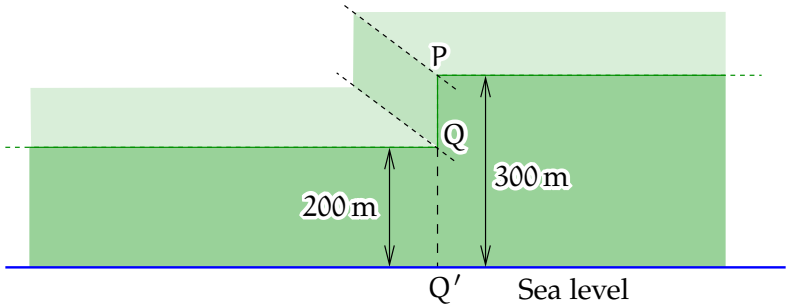


FIGURE 6.7. A special terrain shape. The vertical rock wall at PQ is also straight on a map and extends to infinity in both directions.

attraction formula for the Bouguer plate). Algebraic sign?

2. Calculate the terrain correction at point Q. Algebraic sign?
3. If at point P it is given that the free-air anomaly is 50 mGal, how much is the Bouguer anomaly at the point?
4. If at point Q it is given that the Bouguer anomaly is 22 mGal, how much is the free-air anomaly at the point?

Answers

1. The terrain correction at point P is the change in gravity if the terrain is filled up on the left side up to a level of 300 m. This means *adding half a Bouguer plate*, thickness 100 m, below the level of P. The effect (projected onto the vertical direction) is

$$\begin{aligned} \mathrm{TC} &= \tfrac{1}{2} \cdot 2\pi G\rho \cdot H = \\ &= \tfrac{1}{2} \cdot 0.1119\,{}^{\mathrm{mGal}}/_{\mathrm{m}} \cdot 100\,\mathrm{m} = 5.595\,\mathrm{mGal}. \end{aligned}$$

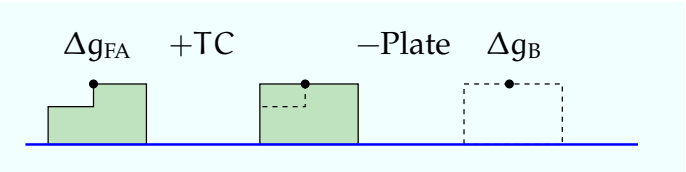
2. The terrain correction at point Q is the change in gravity if we *remove the half Bouguer plate* to the right of and above the point, which is 100 m thick. Its vertical gravity effect is, as calculated above,

$$\mathrm{TC} = 5.595\,\mathrm{mGal},$$

and, because a semi-plate is *removed* that is *above* the level of point Q, the algebraic sign of TC is again positive.

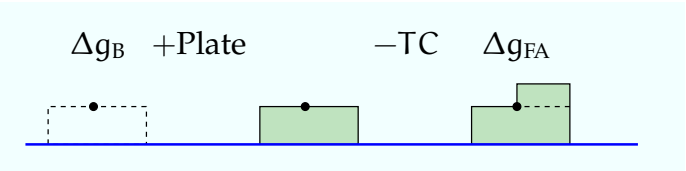
3. Free air to Bouguer:

$\Delta g_{FA}(P)$	50.000 mGal
TC	+5.595 mGal
Bouguer plate removal, 300 m	−33.570 mGal
$\Delta g_{B}(P)$	22.025 mGal



4. Bouguer to free air:

$\Delta g_{B}(Q)$	22.000 mGal
Bouguer plate addition, 200 m	+22.380 mGal
TC "uncorrection"	−5.595 mGal
$\Delta g_{FA}(Q)$	38.785 mGal



6.4 Spherical Bouguer anomalies

More recently, *spherical Bouguer anomalies* have also been calculated, for example Balmino et al. (2012); Kuhn et al. (2009); Hirt and Kuhn (2014). In this calculation, the topography *and bathymetry* of the whole Earth is taken into account, in spherical geometry (the error caused by neglecting the Earth's flattening is in this calculation negligible). Spherical Bouguer anomalies differ from Bouguer-plate anomalies in four ways:

1. The attraction of a Bouguer shell of thickness H is $4\pi G\rho H$, twice as much as the corresponding Bouguer-plate attraction. The

remote part of the shell contributes as much attraction as the neighbourhood of the evaluation point!

2. The bathymetry of the oceans is accounted for[2] by replacing the water with standard-density crustal rock. This contribution to the anomalies is *positive.*
3. The topography and bathymetry of remote parts of the globe are also taken into account realistically. As most of the Earth is covered by deep ocean, this causes a strong *positive general bias*, also in low-lying areas, where the planar Bouguer reduction is typically small.
4. As the terrain correction is now calculated over the whole globe in spherical geometry — though only for the topography. Therefore the rule that all its contributions are positive no longer applies: Abrehdary et al. (2016) report that in places near mountain ranges below the local horizon, the spherical terrain correction can be as negative as −200 mGal.

There is a large systematic difference between the planar and spherical Bouguer anomalies, which however is very long-wavelength in nature, and even in an area the size of Australia almost a constant, −18.6 mGal within a variation interval of a few milligals. The details in the Bouguer maps look the same (Kuhn et al., 2009).

Just for fun, we compute the net mass effect of doing the complete spherical Bouguer reduction globally. The mean height of the land topography is 800 m, land occupying 29 % of the globe. The mean ocean depth is 3700 m, corresponding to an equivalent rock depth to be "filled in" of

$$3700 \times \frac{2670 - 1030}{2670}\,\mathrm{m} = 2272\,\mathrm{m},$$

assuming a density for crustal rock of 2670 $^{\mathrm{kg}}/_{\mathrm{m}^3}$, a sea-water density of 1030 $^{\mathrm{kg}}/_{\mathrm{m}^3}$, and ocean occupying 71 % of the globe. The sum weighted

[2] One can also do so, and often does, in connection with the Bouguer-plate correction.

by area is thus

$$(0.29 \times 800 - 0.71 \times 2272)\ \mathrm{m} = -1381\ \mathrm{m}.$$

Interpretation There is not enough topography to fill all of the oceans, even if we are allowed to compress sea water into standard crustal rock. If we try this bulldozing experiment, we will end up 1381 m short or below the current sea level.

puskutraktori

If, instead, we add standard crustal rock to end up at current sea level — the definition of spherical Bouguer reduction! — we will add to the Earth's attraction as sensed from space an amount of $4\pi G\rho \times 1381\ \mathrm{m} = 309\ \mathrm{mGal}$.

The global mean planar Bouguer reduction, as well as the difference between spherical and planar Bouguer reductions, on average over the globe, will be half of this, $\approx 155\ \mathrm{mGal}$. As the global mean free-air anomaly is zero, the global mean spherical Bouguer anomaly will be 309 mGal, the bulk of the positivity being found over the deep ocean.

6.5 Helmert condensation

[3] An often-used method, proposed by Friedrich Robert Helmert,[3] for removing the effect of the masses exterior to the geoid is *condensation*. In this method, we shift mathematically all the continental masses vertically downwards to mean sea level into a simple mass-density layer

$$\kappa = \rho H,$$

where H is the height of the topography above sea level and ρ its mean matter density. This mass surface density can be interpreted as a column mass integral:

$$\kappa = \rho \int_R^{R+H} dz.$$

[3] Friedrich Robert Helmert (1843–1917) was a German geodesist known for his work on mathematical and statistical geodesy.

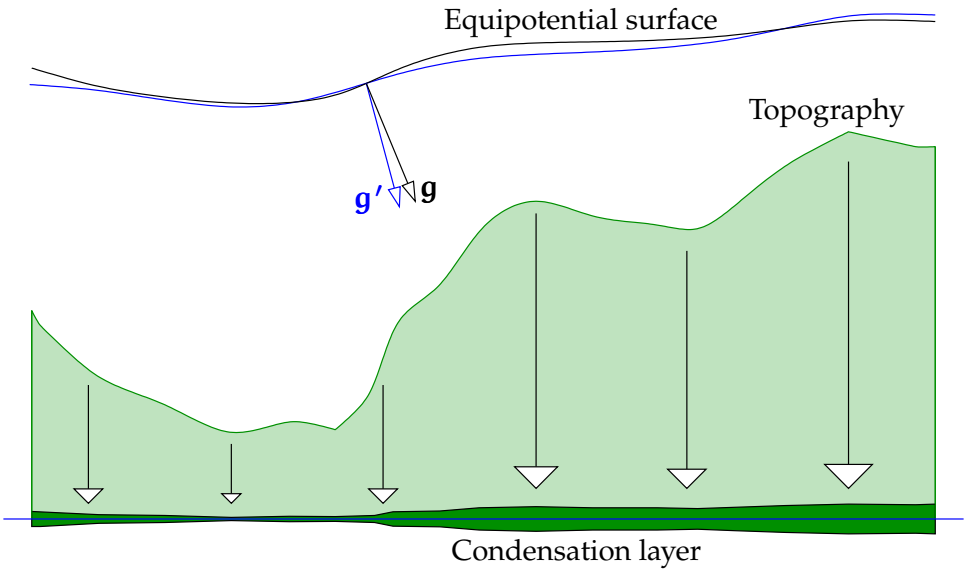


FIGURE 6.8. Helmert condensation and the changes it causes in the gravity field.

For a spherical Earth, the corresponding integral is

$$\kappa = \rho \int_R^{R+H} \left(\frac{r}{R}\right)^2 dr = \rho \frac{1}{R^2} \left[\tfrac{1}{3} r^3\right]_R^{R+H} = \rho H \left(1 + \frac{H}{R} + \tfrac{1}{3}\frac{H^2}{R^2}\right), \quad (6.4)$$

where it is understood that mass is moved from a column cross-section of r^2/R^2 to sea level, where the cross-section is 1.

The advantage of Helmert condensation over Bouguer reduction is that *no mass is being removed*. The Bouguer reduction amounts to the computational removal of topographic masses on a large scale. Therefore, unlike with Bouguer reduction, in Helmert condensation gravity anomalies will not change systematically.

In appendix D we derive series expansions in spherical geometry which express both the exterior and the interior potential as functions of the "degree constituents" of the various powers of the topography $H(\phi, \lambda)$. The extensively presented derivation in the appendix is much-used in gravity field theory to model the gravity effect of the topography. In this theory, issues of convergence are difficult, although we gloss suppeneminen

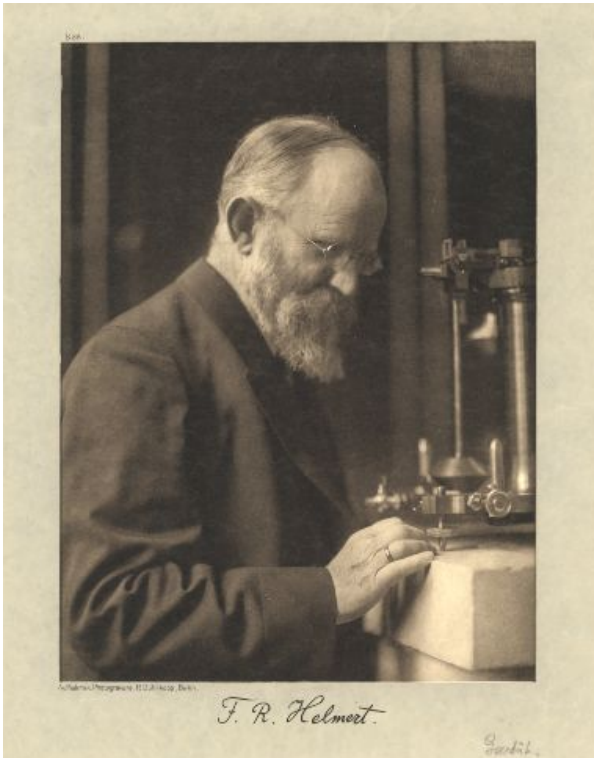


FIGURE 6.9. Friedrich Robert Helmert. Humboldt University Berlin (2017).

over those here.

6.6 Isostasy

6.6.1 *Classical hypotheses*

As early as in the 18th and 19th centuries, also thanks to Bouguer's work in South America as well as that of British geodesists in the Indian Himalayas, it was understood that mountain ranges are not just piles of rock on top of the Earth's crust. The gravity field surrounding the mountains, specifically the deflections of the plumb line, could only be explained by assuming that under every mountain range there was also a "root" made from lighter rock species. The origin of this root was speculated to be the almost hydrostatic behaviour of the Earth's crust

luotiviivan poikkeama

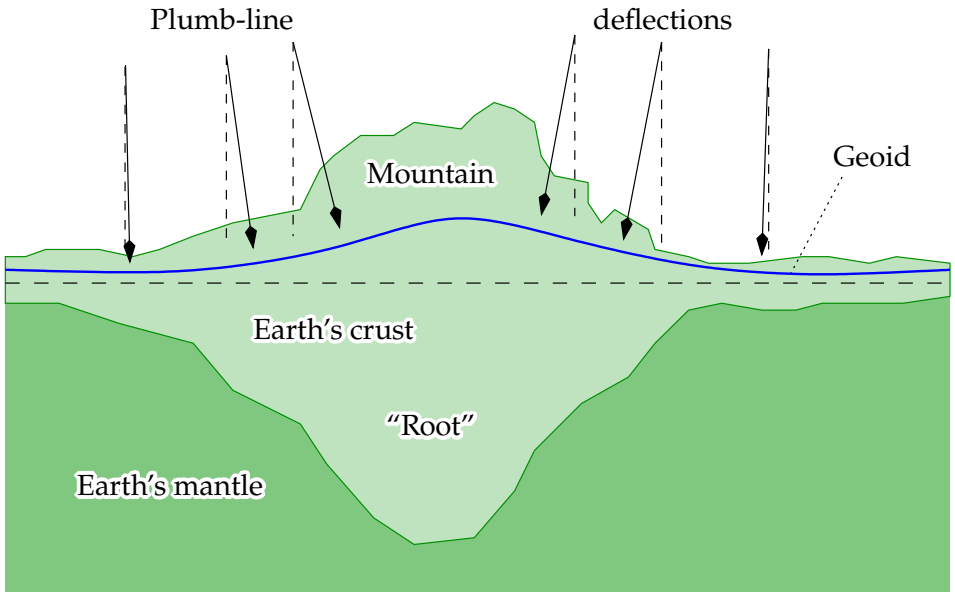


FIGURE 6.10. Isostasy and the deflection of plumb lines towards the mountain.

over geological time-scales. This assumption of hydrostatic equilibrium was called the *hypothesis of isostasy*, also *isostatic compensation*.

Back then, unlike now, it was not yet possible to get a precise or even correct picture using physical methods (seismology) of how these mountain roots are really shaped. That is why simplified working hypotheses were formulated.

One classical isostatic hypothesis is the Pratt–Hayford hypothesis. This was proposed by J. H. Pratt[4] in the middle of the 19th century (Pratt, 1855, 1859, 1864), and J. F. Hayford[5] developed the mathematical tools needed for computation. According to this hypothesis, the matter density of the "root" under a mountain would vary with the height of the mountain, so that under the highest mountains would be the

[4] John Henry Pratt (1809–1871) was a British clergyman and mathematician who worked as the archdeacon of Kolkata, India. Wikipedia, John Pratt.

[5] John Fillmore Hayford (1868–1925) was a United States geodesist who studied isostasy and the figure of the Earth.

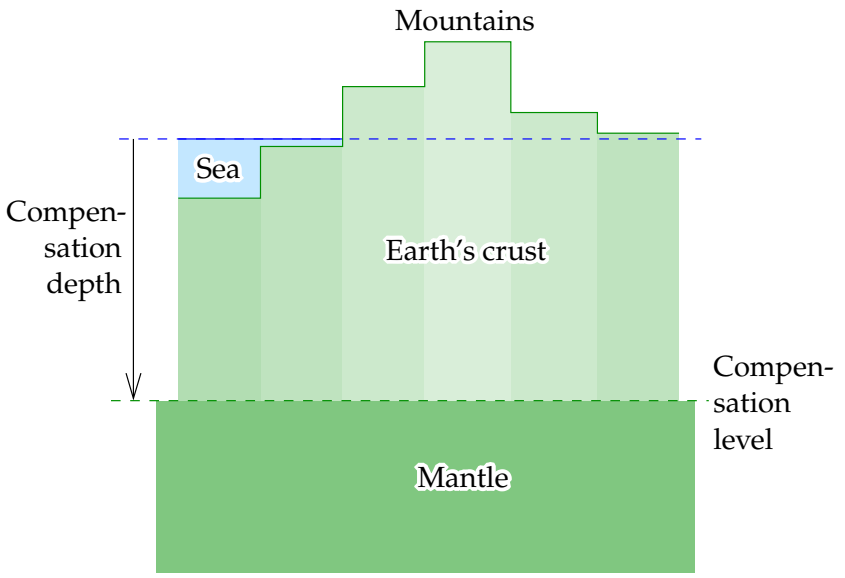


FIGURE 6.11. Pratt–Hayford isostatic hypothesis.

lightest material, and the boundary between this light root material and the denser material of the Earth's mantle would be at a fixed depth. This model, which nowadays finds little acceptance, is illustrated in figure 6.11.

6 Another classical isostatic hypothesis is due to G. B. Airy.[6] Because V.
7 A. Heiskanen[7] used it extensively and developed its mathematical form, it is called the Airy–Heiskanen model. In this model, it is assumed that the mass density of the "root" is fixed, and that the isostatic compensation is realised by varying the depth to which the root extends down into the Earth's mantle. In our current understanding, this corresponds better to what is really happening inside the Earth. This hypothesis is illustrated in figure 6.12.

[6] George Biddell Airy PRS (1801–1892) was an English mathematician and astronomer, "Astronomer Royal" 1835–1881.

[7] Veikko Aleksanteri Heiskanen (1895–1971), "the great Heiskanen" (Hermans, 2007) was a Finnish geodesist who also worked in Ohio, USA. He is known for his work on isostasy and global geoid modelling, the "Columbus geoid". See Kakkuri (2008).

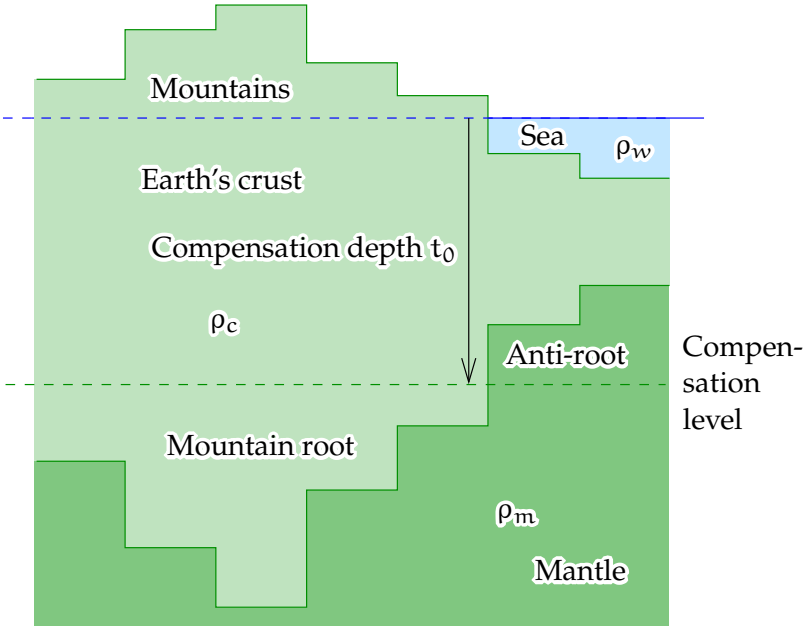


FIGURE 6.12. Airy–Heiskanen isostatic hypothesis.

6.6.2 *Calculation formulas*

Airy's isostatic hypothesis assumes that in every place the total mass of a column of matter is the same. So, let the density of the Earth's crust be

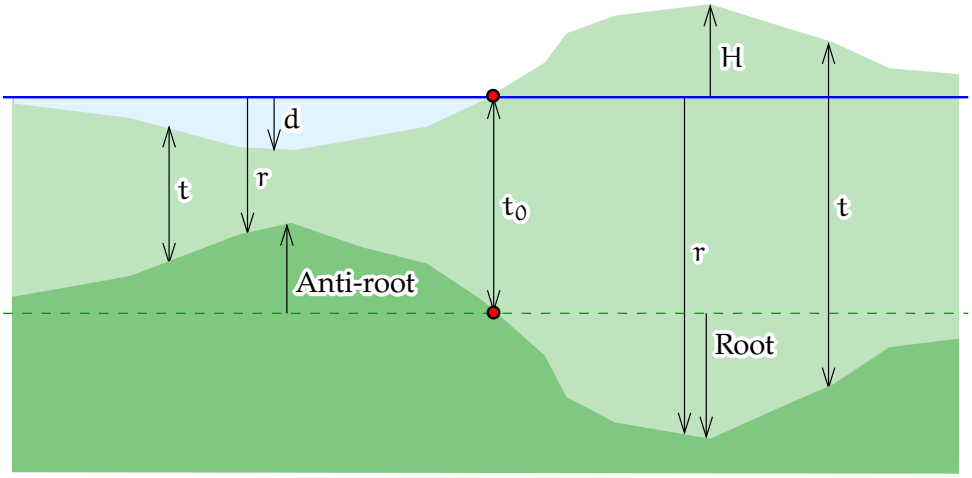


FIGURE 6.13. Quantities in isostatic compensation.

ρ_c, the density of the mantle ρ_m, the density of sea water ρ_w, sea depth d, crustal thickness t, and topographic height H. We obtain

$$t\rho_c + d\rho_w - (t+d)\,\rho_m = c \quad\Longrightarrow\quad t = -\frac{d\,(\rho_m - \rho_w) + c}{\rho_m - \rho_c}$$

on the sea, and

$$t\rho_c - (t-H)\,\rho_m = c \quad\Longrightarrow\quad t = \frac{H\rho_m - c}{\rho_m - \rho_c}$$

[8] on land. c is a suitable constant.[8] Here we have conveniently forgotten about the curvature of the Earth: we use the "flat Earth model".

Under land, the depth of a mountain root is

$$r = t - H = \frac{H\rho_m - c}{\rho_m - \rho_c} - \frac{H\rho_m - h\rho_c}{\rho_m - \rho_c} = \frac{H\rho_c - c}{\rho_m - \rho_c}.$$

Similarly under the sea

$$r = t + d = -\frac{d\,(\rho_m - \rho_w) + c}{\rho_m - \rho_c} + \frac{d\rho_m - d\rho_c}{\rho_m - \rho_c} = -\frac{d\,(\rho_c - \rho_w) + c}{\rho_m - \rho_c}.$$

In the equations, the constant c is, at least from the viewpoint of isostatic equilibrium, arbitrary and expresses the fact that the level from which one computes the depth of the root — less precisely, the "average thickness of the crust" — can be chosen arbitrarily.

Another approach: instead of c, use the "zero topography compensation level", for short *compensation depth*, t_0, to be computed from the above equations by setting $H = d = 0$:

$$t_0\,(\rho_c - \rho_m) = c.$$

This yields under the land the root depth

$$r = \frac{H\rho_c - t_0\,(\rho_c - \rho_m)}{\rho_m - \rho_c} = t_0 + H\frac{\rho_c}{\rho_m - \rho_c}, \tag{6.5}$$

[8] Its dimension, after multiplication with ambient gravity g, is *pressure*: according to Archimedes' law, the pressure of the crustal (plus sea-water) column minus the pressure of the column of displaced mantle material.

and under the sea

$$r = -\frac{d\,(\rho_c - \rho_w) + t_0\,(\rho_c - \rho_m)}{\rho_m - \rho_c} = t_0 - d\frac{\rho_c - \rho_w}{\rho_m - \rho_c}, \qquad (6.6)$$

somewhat simpler equations that are also more intuitive.

Still a third form:

$$H\rho_c + (-r)\,(\rho_m - \rho_c) = c,$$
$$(-d)\,(\rho_c - \rho_w) + (-r)\,(\rho_m - \rho_c) = c.$$

In other words,

$$\sum_{\text{interfaces}} (\text{deviation} \times \text{density contrast}) = \text{constant}.$$

The effect of the different isostatic hypotheses on gravity is pretty much the same: the hypotheses cannot be distinguished based on gravity measurements alone. The effect of the choice of hypothesis on the geoid is stronger.

6.6.3 Example: Norway

The southern Norwegian *Hardanger plateau* (*Hardangervidda*) is a highland at, on average, 1100 m above sea level. It is the largest peneplain in Europe, a national park, and a popular tourist attraction, being traversed by the *Bergensbanen*, the highest regular railway in Northern Europe. puolitasanko

The *Norwegian Sea* is the part of the Atlantic Ocean adjoining Norway, and does not belong to the continental shelf. It is on average 2 km deep. mannerjalusta

Questions

1. What is the depth of the root under the Hardanger plateau, relative to the compensation depth t_0?
2. What is the negative depth of the anti-root under the Norwegian Sea, relative to the same compensation depth?
3. What is the *relative* depth of the root of the Hardanger plateau, compared to the nearby Norwegian Sea?

Answers

1. We use equation 6.5, finding

$$r - t_0 = H\frac{\rho_c}{\rho_m - \rho_c} = \\ = 1100\,\text{m} \times \frac{2670\,\text{kg/m}^3}{(3370 - 2670)\ \text{kg/m}^3} = 4196\,\text{m}.$$

   Here we have used standard densities for crustal and mantle rock, respectively.

2. We use equation 6.6, finding

$$r - t_0 = -d\frac{\rho_c - \rho_w}{\rho_m - \rho_c} = \\ = -2000\,\text{m} \times \frac{(2670 - 1030)\ \text{kg/m}^3}{(3370 - 2670)\ \text{kg/m}^3} = -4686\,\text{m},$$

   using the standard density value for sea water.

3. The depth contrast between root and anti-root is $4196 - (-4686)$ m $= 8882$ m. For perspective, Mount Everest is 8848 m above sea level.

6.6.4 *The modern understanding of isostasy*

Nowadays we have a much better understanding of the internal situation in the Earth. However, isostasy continues to be a valid concept. A more realistic understanding of the internal structure of the Earth is given in figure 6.14.

An important subject for current research is the effect on vertical motion of the Earth's crust of the growing and melting of the ice masses of the Earth, like the continental ice sheets. This includes both the direct effect of the varying ice masses and the effect of the variation of the oceanic water masses. *Paleo-research* concentrates on the changes over the glacial cycle, while modern retreats of glaciers, as in Alaska and on Spitsbergen, cause their own, observable local uplift of the Earth's crust. More in chapter 12.

mannerjäätikkö

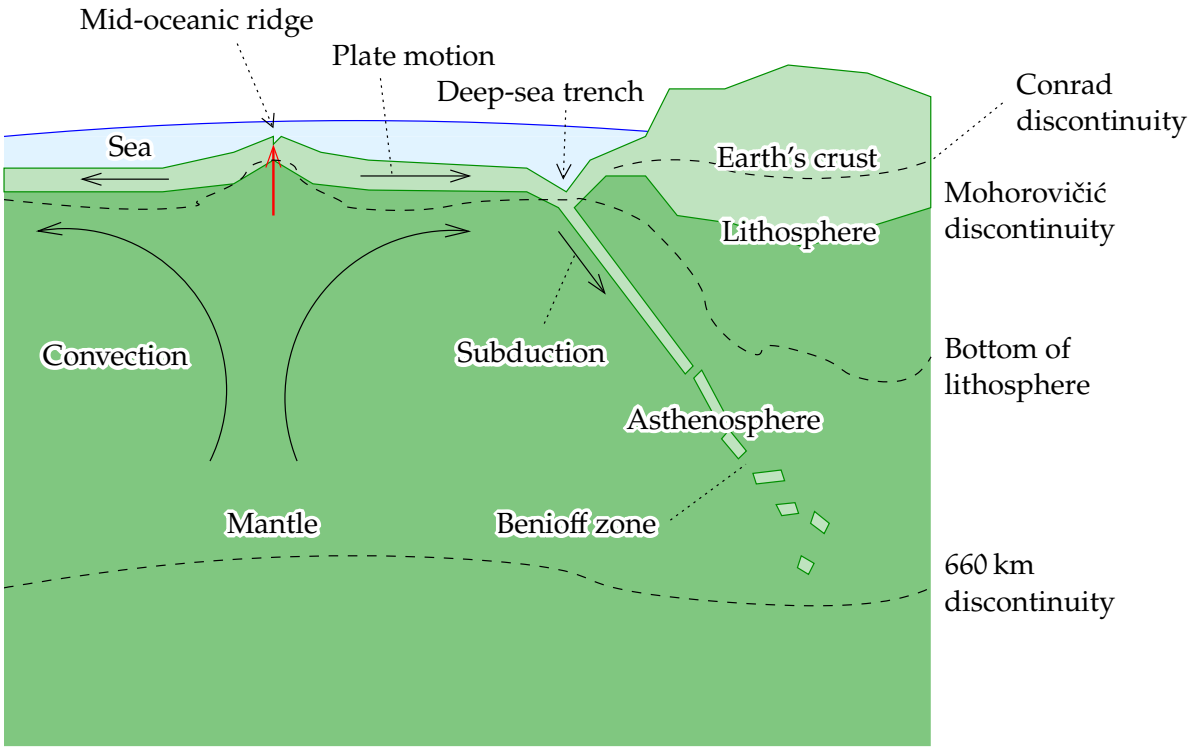


FIGURE 6.14. The modern understanding of isostasy and plate tectonics. Deep-sea trenches are known to be in isostatic disequilibrium.

6.6.5 *Example: Fennoscandian land uplift*

During the last glacial maximum, some 20 000 years ago, Fennoscandia was covered by a continental ice sheet of thickness up to 3 km.

Questions

1. How much was the Earth's surface depressed by this load, assuming isostatic equilibrium?
2. Currently the land is rising in central Fennoscandia, where the ice thickness was greatest, at a rate of $10\,\mathrm{mm/a}$. How long would it take at this rate for the depression to vanish?

Answers

1. We assume for the ice density a value of $920\,\mathrm{kg/m^3}$. With an

upper mantle density of $3370\,\mathrm{kg/m^3}$ — note that it is Earth's mantle material that is being displaced by the ice, the Earth's crust just transmits the load! See figure 12.1a — we find for the depression

$$\Delta H = 3000\,\mathrm{m} \times \frac{920\,\mathrm{kg/m^3}}{3370\,\mathrm{kg/m^3}} = 819\,\mathrm{m}.$$

2. At the rate of $10\,\mathrm{mm/a}$ it will take $819\,\mathrm{m} \big/ 0.01\,\mathrm{m/a} = 81\,900$ years total for the depression to vanish. Part of this uplift has already taken place since the last deglaciation.
   In reality, of course, the rate has decreased substantially, and will continue to decrease, over time.

6.7 Isostatic reductions

The computational removal of both the topography and its isostatic compensation from the measured quantities of the gravity field is called *isostatic reduction*. It serves two purposes:

- By removing as many as possible "superficial" effects from the gravity field, we are left with a field where only the effect of the Earth's deep layers remains. This is useful for geophysical studies.
- These "superficial" effects are also generally very local: in spectral language, very *short-wavelength*. By removing those, we are left with a residual field that is smoother, and that can be interpolated or *predicted* better. This is important especially in areas where there is a paucity of real measurement data, like the oceans, deserts, polar areas, etc.

Isostatic anomalies, free-air anomalies to which isostatic reduction has been applied, are very smooth (like Bouguer anomalies), and their *predictive properties* are good. However, unlike Bouguer anomalies, isostatic anomalies are on average zero. They lack the large bias that makes Bouguer anomalies strongly negative especially in mountainous

areas, section 6.2. This of course is because isostatic reduction is only the *shifting* of masses from one place to another — from mountains to roots beneath the same mountains, the mass deficit of which is pretty precisely the same as the mass of the mountains themselves sticking out above sea level — rather than *removal* of masses, which is what Bouguer reduction does.

The reduction methods used in isostatic calculations are similar to those in other reductions. We will discuss them later: numerical integration in the space domain — grid integration, spherical-cap integration, least-squares collocation (LSC), finite elements, etc. — or in the spectral domain, for example FFT and "Fast Collocation".

pienimmän neliösumman kollokaatio

The question of the *hypothesis* assumed to apply is a more interesting one. Traditionally, the Pratt or Airy hypotheses have been used, developed into quantitative methodologies by Hayford or Heiskanen or Vening Meinesz.[9] A newer approach has been to use real measurement data from *seismic tomography* in order to model the interior structure of the Earth. With real measurement data, if reliable, one should get better results.

9

6.8 The "isostatic geoid"

Let us look at how the "isostatic geoid", more precisely the *co-geoid of isostatic reduction*, is computed. Isostatic reduction is one possible method for computationally removing the masses outside the geoid, in order to formulate a boundary-value problem on the geoid.

We can show (Heiskanen and Moritz, 1967, page 142), that under the continents, the isostatic co-geoid is as much as several metres below the geoid. In other words, the indirect effect (the "restore" step) is of this order. Under the oceans, similarly the isostatic co-geoid is somewhat

[9] Felix Andries Vening Meinesz (1887–1966) was a Dutch geophysicist, geodesist and gravimetrist. He wrote with V. A. Heiskanen the textbook *The Earth and its Gravity Field* (1958).

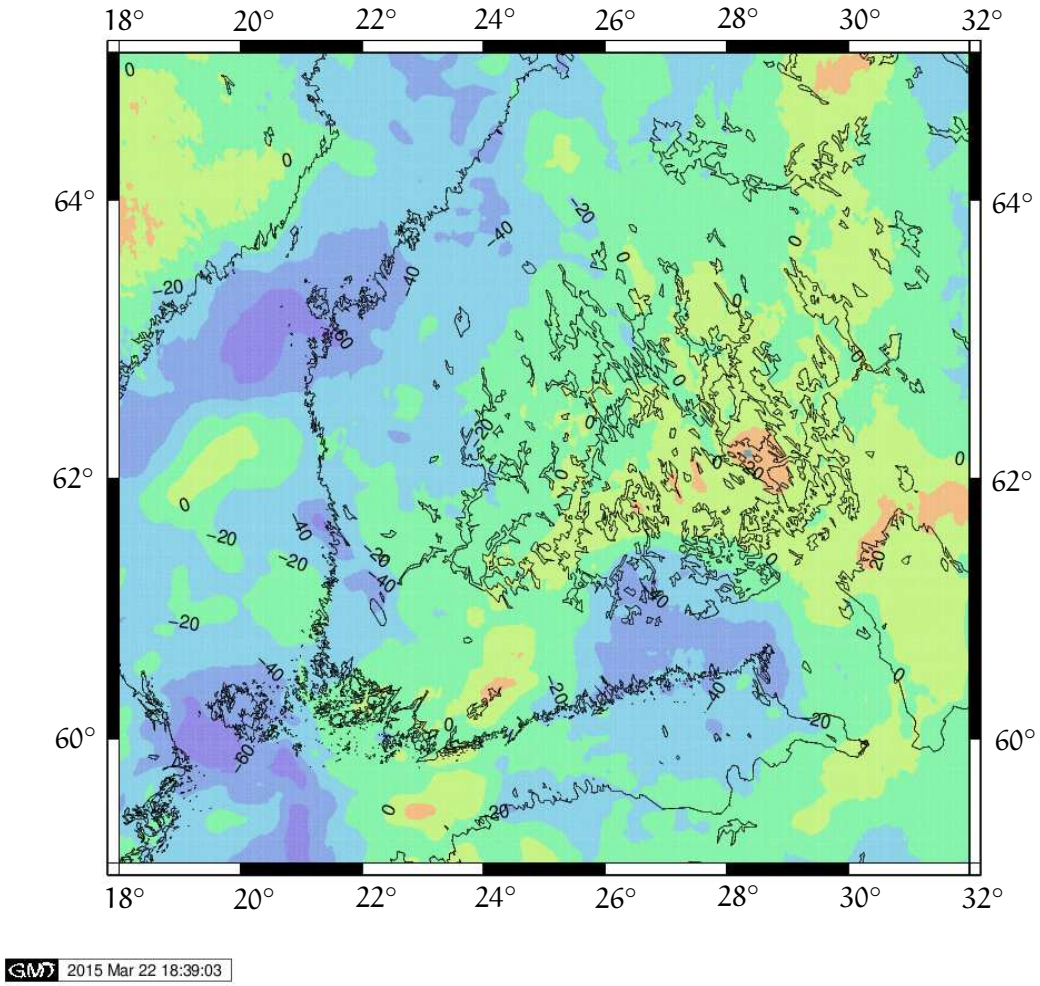


FIGURE 6.15. Isostatic gravity anomalies for Southern Finland. Airy–Heiskanen hypothesis, compensation depth 30 km. Data © Bureau Gravimétrique International (BGI) / International Association of Geodesy, World Gravity Map project. Web service BGI, WGM2012. Here, on the thick, rigid Fennoscandian Shield, the local features of the topography are not isostatically compensated and the map looks rather similar to the free-air anomaly map 5.5 on page 126.

above the geoid.

As one of the requirements for geoid determination methods is a

small indirect effect, it follows that isostatic methods are not perhaps the best possible if the intent is to calculate a model of the geoid or quasi-geoid representing the exterior potential.[10] Heiskanen and Moritz (1967) call on their page 152 the indirect effect "moderate". 10

However, isostatic methods are very suitable for elucidating the interior structure of the Earth, because both the topography and the imprint it makes on the Earth's mantle, the isostatic compensation, are computationally removed.

Research has shown that the great topographic features of the Earth are some 85–90 % isostatically compensated (Heiskanen, 1960). This is valuable information if no other knowledge is available.

This is the second reason why the isostatic geoid is of interest: the gravity field of an Earth from which the effect of mountains has been removed completely — mountain roots and all — can uncover physical unbalances existing in deeper layers, and processes causing these. Such processes include especially convection currents in the Earth's mantle as well as the possible effect of the liquid outer core of the Earth on these currents. Interesting correlations have been found between mantle convection patterns, the global map of the geoid, and the electric current patterns in the core causing the Earth's magnetic field (Wen and Anderson, 1997; Prutkin, 2008; Kogan et al., 1985).

Isostatic reduction consists of two parts:

- computational removal of the topography
- computational removal of the isostatic compensation of the topography.

It is possible to calculate both parts exactly using prism integration, see section 6.3. Here however we shall gain understanding by a qualitative approach. We *approximate* both parts with a single mass-density layer, with density for example $\kappa = \rho H$ for the topography. We place the first

[10] Of course Bouguer reduction is even worse! The indirect effect can be hundreds of metres.

layer at level $H = 0$, and the second, density

$$-\kappa \left(\frac{R}{R-D}\right)^2,$$

at *compensation depth* D. This choice preserves the total mass of the Earth. The situation is depicted — in flat-Earth approximation — in figure 6.16.

In the following we use the "generating function" equation 8.7,

$$\frac{1}{\ell} = \frac{1}{R} \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+1} P_n(\cos\psi),$$

together with the single mass-density layer equation 1.15:

$$V = G \iint_{\text{surface}} \frac{\kappa}{\ell}\, dS = GR^2 \iint_{\sigma} \frac{\kappa}{\ell}\, d\sigma.$$

We obtain for the potential field of the mass-density layer at sea level, when the evaluation point is also placed at sea level, $H = 0 \implies r = R$:

$$V_{\text{top}} = GR \iint_{\sigma} \kappa \sum_{n=0}^{\infty} P_n(\cos\psi)\, d\sigma$$

and with the density layer at compensation depth (source level $R - D$, evaluation level R):

$$V_{\text{comp}} =$$
$$= G(R-D) \iint_{\sigma} \left(-\kappa \left(\frac{R}{R-D}\right)^2\right) \sum_{n=0}^{\infty} \left(\frac{R-D}{R}\right)^{n+1} P_n(\cos\psi)\, d\sigma =$$
$$= -GR \iint_{\sigma} \kappa \sum_{n=0}^{\infty} \left(\frac{R-D}{R}\right)^{n} P_n(\cos\psi)\, d\sigma,$$

from which the combined effect of the reduction is ($n = 0$ drops out):

$$\delta V_{\text{iso}} = -\left(V_{\text{top}} + V_{\text{comp}}\right) =$$
$$= -GR \iint_{\sigma} \kappa \sum_{n=1}^{\infty} \left(1 - \left(\frac{R-D}{R}\right)^{n}\right) P_n(\cos\psi)\, \sigma. \quad (6.7)$$

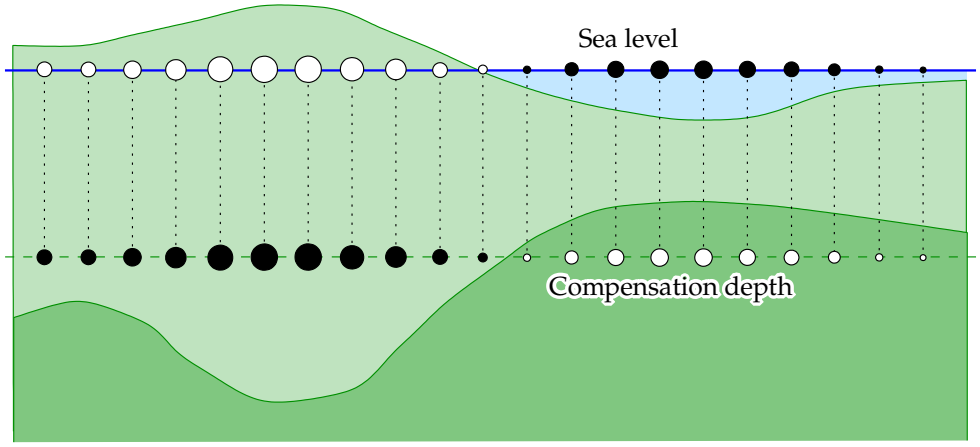


FIGURE 6.16. Isostatic reduction as a pair of surface density layers.

Here, the mass density per unit of surface area κ is

$$\kappa = \begin{cases} \rho_c H & \text{if } H \geqslant 0, \\ (\rho_c - \rho_w)\, H & \text{if } H < 0, \end{cases}$$

so we replace ocean depths with equivalent "dry" depths.[11] Now we use again the degree constituent equation, Heiskanen and Moritz (1967) equation 1-71, or our equation 3.9, in the following form:

[11] asteosuus-yhtälö

$$\kappa_n(\phi,\lambda) \overset{\text{def}}{=} \frac{2n+1}{4\pi} \iint_\sigma \kappa(\phi',\lambda')\, P_n(\cos\psi)\, d\sigma'.$$

Multiplying both sides by the factor

$$-\frac{4\pi G R}{2n+1}\left(1-\left(\frac{R-D}{R}\right)^n\right)$$

and moving it inside the integral, we obtain

$$-\frac{4\pi G R}{2n+1}\left(1-\left(\frac{R-D}{R}\right)^n\right)\kappa_n(\phi,\lambda) =$$
$$= -GR\iint_\sigma \kappa(\phi',\lambda')\left(1-\left(\frac{R-D}{R}\right)^n\right) P_n(\cos\psi)\, d\sigma'.$$

[11] This works on dry land and on the ocean. Lakes, glaciers and areas like the Dead Sea are more complicated.

Summation yields expression 6.7 above:

$$-\sum_{n=1}^{\infty}\frac{4\pi GR}{2n+1}\left(1-\left(\frac{R-D}{R}\right)^n\right)\kappa_n(\phi,\lambda)=$$
$$=-GR\iint_\sigma\kappa(\phi',\lambda')\sum_{n=1}^{\infty}\left(1-\left(\frac{R-D}{R}\right)^n\right)P_n(\cos\psi)\,d\sigma'=\delta V_{iso}.$$

It follows that

$$\delta V_{iso}=-\sum_{n=1}^{\infty}\frac{4\pi GR}{2n+1}\left(1-\left(\frac{R-D}{R}\right)^n\right)\kappa_n(\phi,\lambda)=$$
$$=-\sum_{n=1}^{\infty}\frac{2}{2n+1}R\left(1-\left(\frac{R-D}{R}\right)^n\right)2\pi G\kappa_n=$$
$$=-\sum_{n=1}^{\infty}\frac{2}{2n+1}R\left(1-\left(\frac{R-D}{R}\right)^n\right)(A_B)_n\,.$$

Here we have used the notation $A_B = 2\pi G\kappa$. This represents the equivalent Bouguer-plate attraction of a mass-density layer κ, and its degree constituents are $(A_B)_n = 2\pi G\kappa_n$.

[12] Let us first look at the contribution from[12] $1 < n \leqslant N \overset{\text{def}}{=} {}^{R}/_{D}$. Then, as

$$\left(\frac{R-D}{R}\right)^n \approx 1-\frac{nD}{R},$$

the following approximation holds:

$$\delta V_{iso}\approx-\sum_{n=1}^{N}\frac{2nD}{2n+1}(A_B)_n\approx-\sum_{n=1}^{N}D\,(A_B)_n\approx-D\widetilde{A}_B,$$

[12] The contribution from degree numbers $n > {}^{R}/_{D}$ is

$$\delta V_{iso}\approx-\sum_{n=N+1}^{\infty}\frac{2R}{2n+1}(A_B)_n\,,$$

where the terms are small and rapidly falling to zero. In this degree range, the mass-density layer approximation for the topography and its compensation breaks down, but it hardly matters as these short wavelengths aren't even isostatically compensated.

and

$$\delta N_{iso} = \frac{\delta V_{iso}}{\gamma} \approx -\frac{D\widetilde{A}_B}{\gamma} \approx -\frac{DA_B}{\gamma}. \tag{6.8}$$

This is the *indirect effect of isostatic reduction*.

Let us substitute realistic values. Let the depth of the Mohorovičić[13] discontinuity be on average $\sim 20\,\mathrm{km}$.[14]

On land $H \approx 0.8\,\mathrm{km}$, the Earth's mean topographic height, and we obtain $\delta N_{iso,\,land} \approx -1.8\,\mathrm{m}$.

On the ocean $H \approx -3.7\,\mathrm{km}$ on average. We must still multiply by the ratio

$$\frac{\rho_c - \rho_w}{\rho_c} = \frac{2670 - 1030}{2670},$$

in order to take the water into account. We obtain $\delta N_{iso,\,sea} \approx +5.1\,\mathrm{m}$.

In other words, this effect can be sizeable.

Note, however, that the above calculation used the equivalent Bouguer-plate attraction

$$A_B = \sum_{n=0}^{N} 2\pi G \kappa_n,$$

whereas equation 6.8 contains

$$\widetilde{A}_B = \sum_{n=1}^{N} \frac{2n}{2n+1} 2\pi G \kappa_n \approx \sum_{n=1}^{N} 2\pi G \kappa_n,$$

which lacks the zeroth-degree constituent κ_0. in other words, the global average of $\widetilde{A}_B$, and thus of δN_{iso}, over continents and oceans would have to vanish, because of the assumption made that isostatic reduction does not change the total mass of the Earth. The calculated value is, however,

$$\overline{\delta N_{iso}} = 0.29 \cdot \delta N_{iso,\,land} + 0.71 \cdot \delta N_{iso,\,sea} = 3.1\,\mathrm{m}.$$

[13] Andrija Mohorovičić (1857–1936) was a Croatian meteorologist and a pioneer of modern seismology.

[14] Under the continents, the depth is 35 km, under the oceans 7 km below the sea floor (Encyclopaedia Britannica, Moho).

With this value as a correction, we obtain

$$\delta N_{\text{iso, land}} \approx -1.8\,\text{m} - 3.1\,\text{m} = -4.9\,\text{m},$$
$$\delta N_{\text{iso, sea}} \approx +5.1\,\text{m} - 3.1\,\text{m} = +2.0\,\text{m}.$$

These are values for typical, extended continental or ocean areas, and only indicative. Precise calculation needs to be numerical.

Equation 6.8 is, through the Bouguer-plate attraction A_B, *linear* in the height H. This means that every added kilometre of topography causes about $-2.2\,\text{m}$ in the quantity $\delta N_{\text{iso, land}}$, and every added kilometre of bathymetry similarly $+1.4\,\text{m}$ in the quantity $\delta N_{\text{iso, sea}}$. We may also conclude that in the isostatic reduction's effect on the geoid – at least for longer wavelengths $2\pi R/n$, longer than the compensation depth D — all wavelengths are represented in the spectrum in approximately in the same proportions as in the topography itself, and the effect is in fact proportional to the topography.

Self-test questions

1. Which effects are computationally removed in
   (a) the simple Bouguer reduction?
   (b) the terrain-corrected Bouguer reduction?
   (c) the isostatic reduction?
2. Why is the terrain correction always positive?
3. Why do Bouguer anomalies have good interpolation properties, and on what condition — in other words, which additional information must be available at the interpolation stage?
4. How was it discovered that mountains have roots?
5. Explain the isostatic hypotheses of Pratt–Hayford and Airy–Heiskanen.

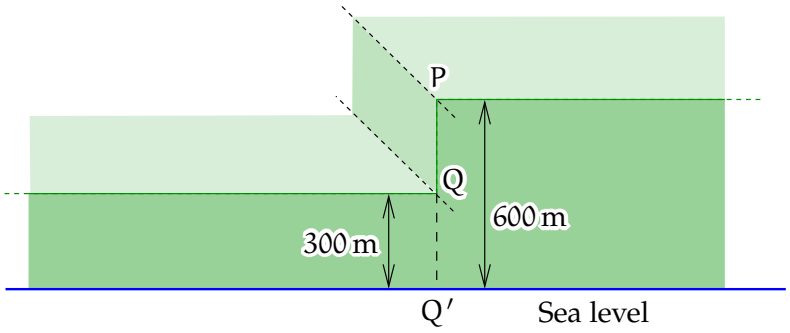


FIGURE 6.17. Terrain shape.

Exercise 6−1: Gravity anomaly

Given point P, height above sea level $H = 500\,\text{m}$, local gravity $g_P = 9.82\,\text{m/s}^2$. Normal gravity at sea level for local latitude φ is $\gamma_0(\varphi) = 9.820\,192\,\text{m/s}^2$.

1. Compute point P's free-air anomaly Δg.
2. Compute point P's Bouguer anomaly (without terrain correction) Δg_B.

Exercise 6−2: Bouguer reduction

1. Point P is 500 m above sea level. Its free-air anomaly is $\Delta g_{FA} = 25\,\text{mGal}$. Calculate the Bouguer anomaly Δg_B of the point. Forget about the terrain correction.
2. See section 6.2: Bouguer anomalies. Derive equations 6.2 and 6.3 anew, assuming that the mean density of the Earth's crust is $\rho = 3370\,\text{kg/m}^3$.

Exercise 6−3: Terrain correction and Bouguer reduction

Given the terrain shape, figure 6.17.

The vertical rock wall PQ is also straight on a map and extends in both directions ("into" and "out of" the paper) to infinity.

Height differences: $PQ' = 600\,\text{m}$, $QQ' = 300\,\text{m}$.

1. Compute the *terrain correction* at point P.

   Hint: use the equation for the attraction of a Bouguer plate. Here is a half Bouguer plate, with only half the attraction of a full one.
2. Compute the terrain correction at point Q. *What is the algebraic sign?*
3. If at point P it is given that the free-air anomaly is 60 mGal, how much is the Bouguer anomaly at the point? (Use the complete Bouguer reduction.)
4. If it is given that at point Q the Bouguer anomaly is 10 mGal, how much is the point's free-air anomaly?

Exercise 6−4: Isostasy

Assume Airy–Heiskanen isostatic compensation (figure 6.12). The density of the Earth's crust $\rho_c = 2670\,\text{kg/m}^3$, density of the mantle $\rho_m = 3370\,\text{kg/m}^3$, so the density contrast at the crust-mantle interface is
vertaustaso $700\,\text{kg/m}^3$. Let the *reference level* for the interface corresponding to zero topography be $-25\,\text{km}$, so $t_0 = 25\,\text{km}$.

1. Calculate the depth of the "root" of an 8 km high mountain below the reference level −25 km, assuming it is isostatically compensated.
2. The volcano Mauna Kea, Hawaii, is 4 km above sea level, however the surrounding sea is 5 km deep. How deep is the root of Mauna Kea below the reference level?
3. How much is the "anti-root" of the surrounding sea *above* the reference level? Let the density of sea water be $1030\,\text{kg/m}^3$.
4. So, how deep is the root of Mauna Kea *relative to its surroundings*?

Vertical reference systems

7

7.1 Levelling, orthometric heights and the geoid

Heights have traditionally been determined by *levelling*. Levelling is a technique for determining height differences using a level (levelling instrument) and two rods or staffs. The level comprises a telescope and a spirit level, and in the measurement situation the telescope's optical axis, the sight axis, is pointing along the local horizon. Levelling staffs are placed on two measurement points, and through the measuring telescope, measurement values are read off them. The difference between the two values gives the height difference between the two points in metres.

The distance between level and staffs is 40–70 m, as longer distances would cause too large errors due to atmospheric refraction. Longer distances are measured by repeat measurements using several instrument stations and intermediate points.

The height differences ΔH thus obtained are not, however, directly useable. The "height difference" between two points P and Q, obtained by directly summing the height differences ΔH, depends namely also on the path chosen when levelling from P to Q. Also the sum of height differences $\sum_{\bigcirc} \Delta H$ around a closed path *is* (generally) *not zero*.

Geometric height is not a conservative field.

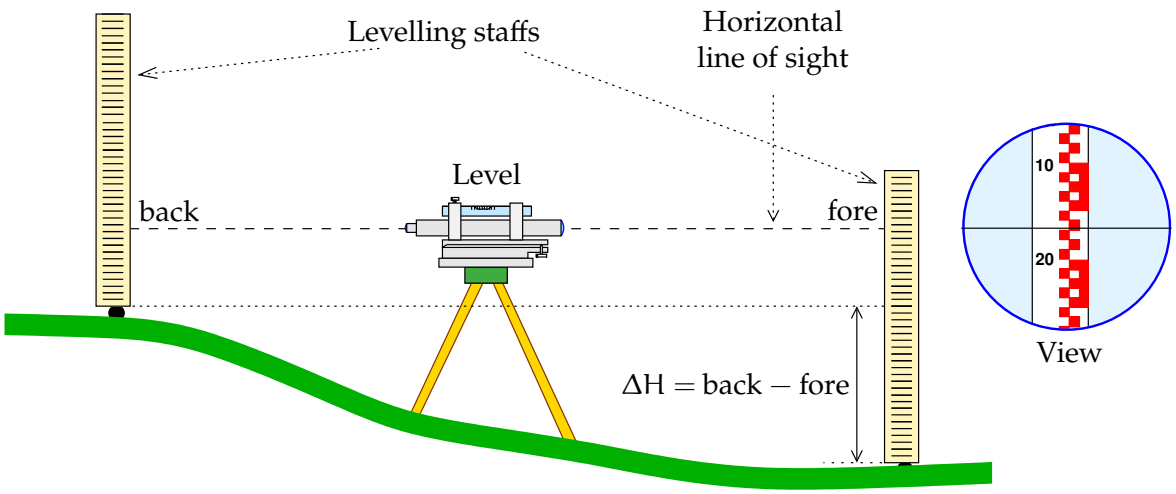


FIGURE 7.1. The principle of levelling.

This is why, in precise levelling, the height differences are always converted to *potential differences*: $\Delta W = -\Delta H \cdot g$, in which g is the local gravity, which is either measured or — like in Finland — interpolated from an existing gravity survey data base. The sum of potential differences around a closed loop *is always zero*: $\sum_{\bigcirc} \Delta W = 0$.

For the potential at an arbitrary terrain point P we find

$$W_P = W_0 - \sum_{\text{sea level}}^{P} (\Delta H \cdot g)\,,$$

the summation being done directly from sea level (potential W_0) up to point P. The quantity

$$C_P = -(W_P - W_0) = \sum_{\text{sea level}}^{P} (\Delta H \cdot g)\,,$$

which is positive above sea level, is called the *geopotential number* of point P.

korkeus-vertaustaso

W_0 is the potential of the national height reference level. In Finland, the reference level of the old N60 system is in principle the mean sea level in Helsinki harbour at the beginning of 1960, which is why the

system is called N60. However, the precise realisation is a special pillar in the garden of the Helsinki astronomical observatory in Kaivopuisto.[1] The new Finnish height system is called N2000, and the realisation of its reference level is a pillar at the Metsähovi research station. In practice N2000 heights are, at the decimetric precision level, heights over the Amsterdam NAP datum.

Other countries have their own, similar height reference or datum points: Russia has Kronstadt, Western Europe the widely used Amsterdam datum NAP, southern Europe has the old Austro-Hungarian harbour city of Trieste, North America the North American Vertical Datum 1988 (NAVD88) with the datum point Father Point (Pointe-au-Père)[2] in Rimouski, Quebec, Canada, etc.

7.2 Orthometric heights

To create a vertical reference, it would be simplest to use the original geopotential differences from sea level, the geopotential numbers defined above, $C = -(W - W_0)$, directly as height values. However, this is psychologically and practically difficult: people want their heights to be in metres.

Geopotential numbers have clear advantages: they represent the *amount of energy* that is needed (for a unit test mass) to move to the point from the reference level. Fluids — sea water, but also air, or, on geological time-scales, even bedrock! — always flow downwards and seek the state of minimum energy.

---

[1] However, the value engraved on the pillar is the reference height of the still older NN system, not N60. The correct reference value for N60 for this pillar, 30.513 76 m, is given in the publication Kääriäinen (1966), page 49.

[2] The district Pointe-au-Père of the city of Rimouski was named after the Jesuit priest Father Henri Nouvel (1621?–1701?), who served forty years with the native population of New France. Pointe-au-Père is also notorious as the location of the RMS *Empress of Ireland* shipwreck in 1914, in which over a thousand passengers perished.

FIGURE 7.2. Height reference benchmark in the garden of the Helsinki astronomical observatory in Kaivopuisto, Kääriäinen (1966). Text:

Suomen	*Utgångspunkt för*
tarkka-	*precisionsnivellementet*
vaakituksen	*i Finland*
pääkiintopiste	*30,4652 m öfver noll*
30,4652 m yli nollan	

(Reference benchmark of the precise levelling of Finland, 30.4652 m above zero).

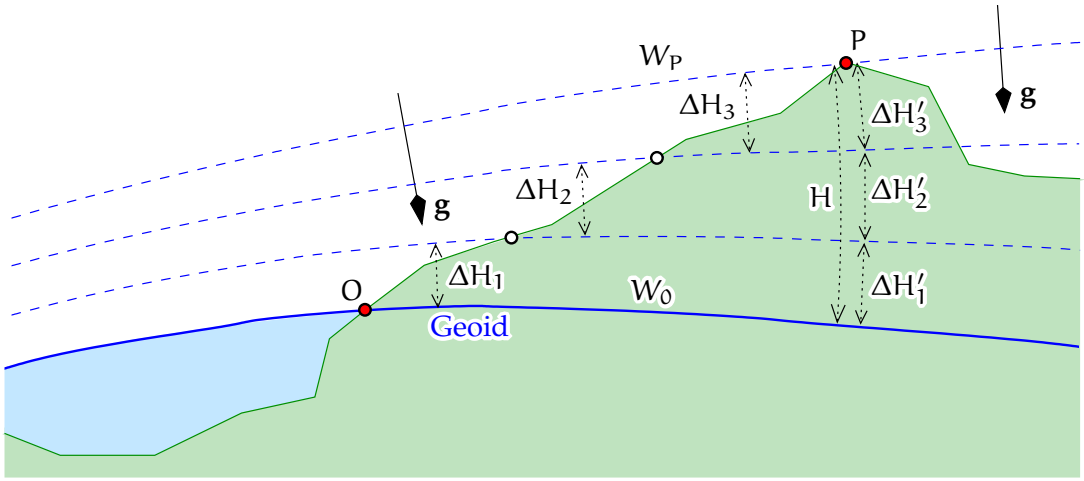


FIGURE 7.3. Levelled heights and geopotential numbers. The height obtained by summing the levelled height differences, $\sum_{i=1}^{3} \Delta H_i$, is not the correct height above the geoid: $\sum_{i=1}^{3} \Delta H_i'$ computed along the plumb line.

The equipotential or *level surfaces* of the geopotential are not parallel: because of this, a journey along the Earth's surface may well go "upwards", to increasing heights above the geoid, although the geopotential number decreases. Thus, water may flow "upwards".

The gravity vector $\mathbf{g}$ is everywhere perpendicular to the level surfaces, and its length is inversely proportional to the distance separating the surfaces.

In Finland, as in many other countries, *orthometric heights* have been long in use. They are physically defined heights above "mean sea level" or the *geoid*. See figure 7.3.

The classical *geoid* is defined as

> *"The level surface of the Earth's gravity field that fits on average best to the mean sea level."*

The orthometric height H of point P is defined as the height obtained by measuring the *distance of* P *from the geoid along the plumb line*. luotiviiva

This is a very physical definition, however not a very operational one, because one (generally) do not get to measure along a plumb line inside the Earth, and the geoid is not even visible there. This is why orthometric heights are calculated from geopotential numbers: if the geopotential number of point P is C_P, we calculate the orthometric height using the equation

$$H = \frac{C_P}{\overline{g}},$$

where $\overline{g}$, *the average gravity along the plumb line*, is

$$\overline{g} = \frac{1}{H} \int_0^H g(z)\, dz,$$

and z is the measured distance from the geoid along the plumb line. Because the equation for $\overline{g}$ already itself contains H, we obtain the solution iteratively, starting from a crude initial estimate for H. The iteration converges fast.

suppeneminen

We shall see that determining precise orthometric heights is challenging, especially in the mountains.

7.3 Normal heights

In Finland, currently, with the N2000 height system, *normal heights* are used. They are, like orthometric heights, heights above mean sea level. The mathematical representation of mean sea level in this case is the *quasi-geoid*. In sea areas, the quasi-geoid is identical to the geoid. Over land, it differs a little from the geoid, and in mountainous areas the difference may be substantial.

7.3.1 *Molodensky's theory*

The renowned theorist M. S. Molodensky (figure 7.5) developed a theory in which the height of a point from "mean sea level" is defined by the following equation:

$$H^* \overset{\text{def}}{=} \frac{C}{\overline{\gamma}_{0H}},$$

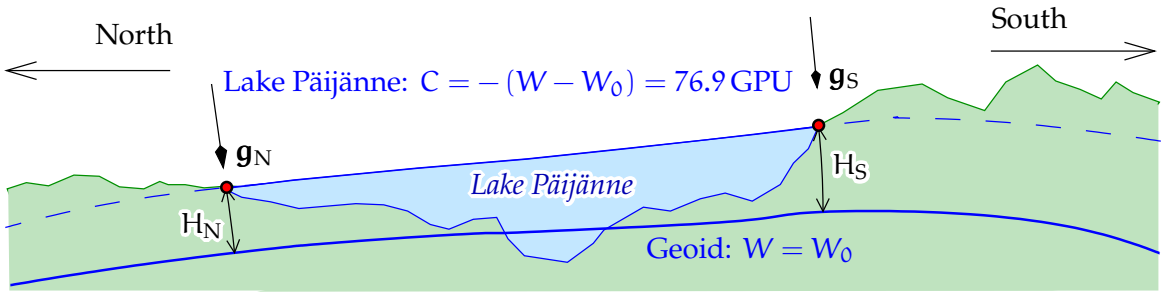


FIGURE 7.4. In terms of orthometric heights, water may sometimes flow "upwards". Although the north and south ends of Lake Päijänne are on the same geopotential level — 76.9 geopotential units below that of mean sea level — the orthometric height of the south end H_S is greater than that of the north end H_N, because local gravity g is stronger in the north than in the south. The height difference in the case of Lake Päijänne is 8 mm (Jaakko Mäkinen, personal communication). Calculation using the normal gravity field yields 6 mm. The balance of 2 mm comes from the difference between the gravity anomalies at the northern and southern ends.

in which $\overline{\gamma_{0H}}$ is the average normal gravity computed between the zero level (reference ellipsoid) and H^* along the *ellipsoidal normal*. So, the method of computing is the same as in the case of orthometric heights, but using the *normal gravity field* instead of the true gravity field.

vertaus-ellipsoidi

Heights "above sea level" are for practical reasons given in metres. For large, continental networks we want to give heights above a computational reference ellipsoid in metres, and thus heights above "sea level" also have to be in metres.

Molodensky also proposed that instead of the geoid, *height anomalies* would be used, the definition of which is

$$\zeta \overset{\text{def}}{=} \frac{T}{\overline{\gamma_{Hh}}}, \tag{7.1}$$

in which now $\overline{\gamma_{Hh}}$ is the average normal gravity at terrain level. More precisely, the average of normal gravity along the ellipsoidal normal over the interval $z \in [H^*, h]$, in which H^* is the normal height of the

FIGURE 7.5. Mikhail Sergeevich Molodensky (1909–1991), source obscure. More photographs and background information in Brovar et al. (2000).

point and h its height from the reference ellipsoid. The parameter z is the distance from the reference ellipsoid reckoned along the ellipsoidal normal. T is the disturbing potential at the point.

häiriö-potentiaali

Based on these assumptions, Molodensky showed that

$$H^* + \zeta = h.$$

This equation is very similar to the corresponding one for orthometric heights and geoid heights

$$H + N = h.$$

Also otherwise ζ, the *height anomaly*, also called "quasi-geoid height", is very close to N, and correspondingly H^* is close to H.

7.3.2 *Molodensky's realisation*

The Molodensky school realised that, because normal gravity along the plumb line is very close to a linear function of place, one could define a

height type that can be computed directly from geopotential numbers, and that also would be compatible with similarly defined, so-called height anomalies, and with geometric heights h reckoned from the reference ellipsoid.

The geometric height h from the reference ellipsoid may be connected to the potential U of the normal gravity field through the following integral equation:

$$U = U_0 - \int_0^h \gamma(z)\, dz.$$

Here, U is the normal potential and γ normal gravity. One level surface of U, $U = U_0$, is also the reference ellipsoid. The variable z is the distance from the ellipsoid along its local normal.[3]

By defining

$$\overline{\gamma_{0h}} \overset{\text{def}}{=} \frac{1}{h} \int_0^h \gamma(z)\, dz \tag{7.2}$$

we obtain

$$h = -\frac{U - U_0}{\overline{\gamma_{0h}}}.$$

By using $W = U + T$ and dividing by $\overline{\gamma_{0h}}$ we obtain

$$\frac{W - W_0}{\overline{\gamma_{0h}}} = \frac{T}{\overline{\gamma_{0h}}} - h$$

assuming $W_0 = U_0$, the normal potential on the reference ellipsoid.

Next, one could define

$$H^+ \overset{?}{=} -\frac{W - W_0}{\overline{\gamma_{0h}}}$$

as a new height type, and

$$N^+ \overset{?}{=} h - H^+ = \frac{T}{\overline{\gamma_{0h}}}$$

as the corresponding new geoid height type. It has however the aesthetic flaw that we divide here by the average normal gravity computed

[3] Here we ignore that the normal gravity vector $\boldsymbol{\gamma}(z)$ is for $z \neq 0$ not precisely parallel with the ellipsoidal normal: the curvature of the field lines of the normal gravity field or normal plumb lines, section 4.3.2.

between the levels 0 and h. This quantity is not operational without a means of determining the ellipsoidal height h.

This suggests the following improvement based on the circumstance that $\gamma(z)$ is a nearly linear function. This means that the vertical derivative $\frac{d}{dz}\gamma$ is nearly constant in the height interval considered.

We define in addition to equation 7.2:

$$\overline{\gamma_{0H}} \stackrel{\text{def}}{=} \frac{1}{H^+}\int_0^{H^+}\gamma(z)\,dz, \qquad \overline{\gamma_{Hh}} \stackrel{\text{def}}{=} \frac{1}{N^+}\int_{H^+}^{h}\gamma(z)\,dz.$$

Now

$$\overline{\gamma_{0H}} \approx \overline{\gamma_{0h}} - \tfrac{1}{2}N^+\frac{d\gamma}{dz} \approx \overline{\gamma_{0h}}\left(1+\frac{N^+}{R}\right), \tag{7.3}$$

$$\overline{\gamma_{Hh}} \approx \overline{\gamma_{0h}} + \tfrac{1}{2}H^+\frac{d\gamma}{dz} \approx \overline{\gamma_{0h}}\left(1-\frac{H^+}{R}\right). \tag{7.4}$$

R is the Earth's radius in spherical approximation: $\frac{d}{dz}\gamma \approx \frac{d}{dr}\gamma \approx -2\gamma/R$.

Next, we also exploit that both N^+/R and H^+/R are $\ll 1$, so

$$\left(1+\frac{N^+}{R}\right)^{-1} \approx \left(1-\frac{N^+}{R}\right), \qquad \left(1-\frac{H^+}{R}\right)^{-1} \approx \left(1+\frac{H^+}{R}\right),$$

and with equations 7.3, 7.4, and the definitions above of H^+ and N^+,

$$\zeta \stackrel{\text{def}}{=} \frac{T}{\overline{\gamma_{Hh}}} = \frac{T}{\overline{\gamma_{0h}}}\cdot\frac{\overline{\gamma_{0h}}}{\overline{\gamma_{Hh}}} \approx N^+\left(1+\frac{H^+}{R}\right) = N^+ + \frac{N^+H^+}{R},$$

$$H^* \stackrel{\text{def}}{=} -\frac{W-W_0}{\overline{\gamma_{0H}}} = -\frac{W-W_0}{\overline{\gamma_{0h}}}\cdot\frac{\overline{\gamma_{0h}}}{\overline{\gamma_{0H}}} \approx H^+\left(1-\frac{N^+}{R}\right) = H^+ - \frac{N^+H^+}{R}.$$

Because the, already small, correction terms N^+H^+/R cancel, we finally obtain

$$H^* + \zeta = H^+ + N^+ = h. \tag{7.5}$$

The quantity $\overline{\gamma_{0H}}$, and thus also normal height H^*, can be, unlike $\overline{\gamma_{0h}}$, computed using *only information obtained by* (spirit or trigonometric)

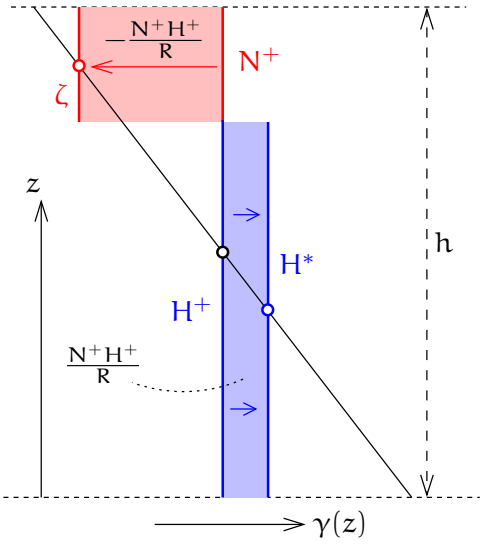


FIGURE 7.6. A graphic cartoon of the proof of Molodensky's realisation. The blue and red areas, which are equal, represent the correction terms which convert N^+ to ζ and H^+ to H^*, respectively. The red and blue arrows stand for the conversion process. The balls represent midpoints of averaging intervals for the function $\gamma(z)$.

levelling, without having to know the height h above the reference ellipsoid, which would again require knowledge of the local geoid.

This was Molodensky's realisation (Molodensky et al., 1962) as early as in 1945, long before the Global Positioning System GPS, or a global, geocentric reference ellipsoid, existed. Back then, continental triangulation networks, like the one of the Soviet Union, were computed on their own, regionally defined reference ellipsoids.

The size of the correction term N^+H^+/R is, for heights of the global geoid up to 110 m, 17 mm for each kilometre of terrain height. The errors remaining after applying this term are microscopically small, because normal gravity is, unlike true gravity, *extremely linear along the plumb line* — as equations 7.3 and 7.4 already assumed.

Figure 7.6 attempts to visualise the derivation.

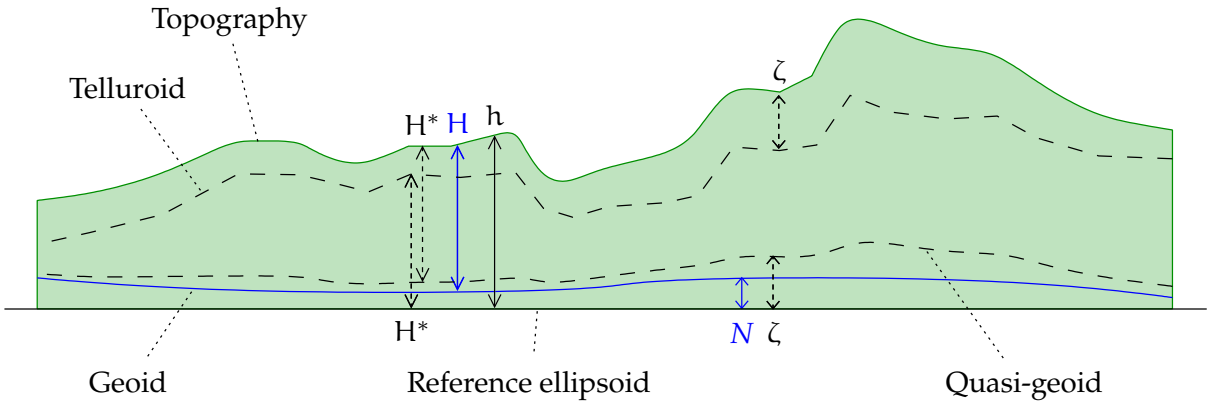


FIGURE 7.7. Geoid, quasi-geoid, telluroid and topography. Note the correlation between the quasi-geoid and topography. Depicted is an area where $N > 0$. The separation between geoid and quasi-geoid is exaggerated.

7.3.3 *Normal height and height anomaly*

Normal height

$$H^* = \frac{C}{\overline{\gamma}} = -\frac{W - W_0}{\overline{\gamma}}, \tag{7.6}$$

in which (recursive definition!)

$$\overline{\gamma} = \overline{\gamma_{0H}} = \frac{1}{H^*}\int_0^{H^*} \gamma(z)\,dz.$$

Height anomaly

$$\zeta = \frac{W - U}{\overline{\gamma_{Hh}}} = \frac{T}{\overline{\gamma_{Hh}}},$$

in which

$$\overline{\gamma_{Hh}} = \frac{1}{\zeta}\int_{H^*}^{h} \gamma(z)\,dz.$$

The height anomaly ζ, which otherwise is a quantity similar to the geoid height N, is however located at the level of the *topography*, not at sea level. The surface formed by points which are a distance H^* above the reference ellipsoid (and thus a distance ζ below or $-\zeta$ above the topography), is called the *telluroid*. It is a mapping

of sorts of the topographic surface: the set of points Q whose *normal* potential U_Q is the same as the *true* potential W_P of the corresponding point P on the topography. See figure 5.4.

Often, as a concession to old habits, we construct a surface that is at a distance ζ above or a distance $-\zeta$ below the reference ellipsoid. This surface is called the *quasi-geoid*. It lacks physical meaning: it is not a level surface, although out at sea it coincides with the geoid. Its short-wave features, unlike those of the geoid, *correlate* with the short-wavelength features of the topography.

Height above the ellipsoid (assumed $U_0 = W_0$)

$$h = \frac{U - U_0}{\overline{\gamma_{0h}}},$$

where

$$\overline{\gamma_{0h}} = \frac{1}{h}\int_0^h \gamma(z)\,dz.$$

The *relationship* between the three quantities is

$$h = H^* + \zeta.$$

In all three cases, the quantity is defined by dividing the potential difference by some sort of "average normal gravity", suitably computed along a segment of the local plumb line. In the case of the height anomaly ζ, a piece of plumb line is used high up, *close to the topographic surface*, between level H^* (telluroid) and level h (topography).

7.4 Difference between geoid height and height anomaly

Normal heights are very *operational*. They are always used together with "quasi-geoid" heights — more correctly: height anomalies — ζ. Orthometric heights — for example Helmert heights — on the other hand are always used together with geoid heights N. For computing

both, H and N, one needs the topographic mass density ρ, for which a standard constant value is often assumed ($2670\,^{kg}/_{m^3}$), and the local vertical gradient of gravity, for which generally the vertical normal gravity gradient ($-0.3084\,^{mGal}/_{m}$) is assumed.

The difference between height anomaly and geoid height is calculated as follows.

1. First, calculate the separation between the quasi-geoid and the "free-air geoid". The free-air geoid is an equipotential surface of the harmonically downwards continued exterior potential. If T_{FA} is the disturbing potential of the exterior, harmonically downwards continued field, then its difference between topography level and sea level is:

$$T_{FA}(H) - T_{FA}(0) = \int_0^H \frac{dT_{FA}(z)}{dz} dz \approx -\Delta g_{FA} H, \tag{7.7}$$

and by using the Bruns equation twice, $\zeta = {}^{T(H)}/_{\gamma} = {}^{T_{FA}(H)}/_{\gamma}$ (height anomaly or quasi-geoid height) and $N_{FA} = {}^{T_{FA}(0)}/_{\gamma}$ ("free-air geoid height", FA = Free Air), we obtain[4]

$$\zeta - N_{FA} \approx -\frac{\Delta g_{FA} H}{\gamma}. \tag{7.8}$$

2. Thus we have obtained the difference between the height anomalies and heights of the free-air geoid. It remains to determine the separation between the free-air geoid and the geoid.

   Let us approximate the topography by a Bouguer plate. Then

   - In the case of the free-air geoid N_{FA} the thickness of the plate is the height H of point P. This is because the free-air geoid is based on the harmonically downwards continued exterior field, meaning that the Bouguer-plate attraction acting at P must also be continued downwards, in other words, taken fully into account.

[4] Here we made the approximation that γ is the same on the topography level as at sea level.

Because the surface mass density of the plate is $H\rho$, its *assumed* attraction is everywhere on the plumb line of point P:

$$2\pi G\rho\, H. \tag{7.9}$$

- Now in the case of the *geoid height* $N = T(0)/\gamma$, we have to be physically realistic: in an arbitrary location z on the plumb line of point P, the Bouguer plate is partly below the location, and partly above it. The attraction is then only

$$2\pi G\rho\, z - 2\pi G\rho\,(H - z) = 2\pi G\rho\,(2z - H)\,. \tag{7.10}$$

By integrating the difference between equations 7.9 and 7.10, like we did for equation 7.7, we obtain

$$\begin{aligned} T(0) - T_{FA}(0) &= 2\pi G\rho \int_0^H \big((2z - H) - H\big)\, dz = \\ &= 2\pi G\rho \left[z^2 - 2Hz\right]_{z=0}^{z=H} = -2\pi G\rho H^2 = -A_B H, \end{aligned}$$

in which A_B is the attraction of a Bouguer plate of thickness H. We obtain again by dividing the equation by normal gravity:

$$N - N_{FA} = -\frac{A_B H}{\gamma}.$$

By subtracting this latest result from equation 7.8, we find

$$\zeta - N = \frac{(-\Delta g_{FA} + A_B)\, H}{\gamma} = -\frac{\Delta g_B H}{\gamma}. \tag{7.11}$$

See also Heiskanen and Moritz (1967), pages 327–328. As the Bouguer anomaly Δg_B is strongly negative in the mountains, it follows that the quasi-geoid is there always *above* the geoid: approximately, using equation 6.2:

$$\zeta - N \approx \frac{0.1119\,\text{mGal/m}}{9.81\,\text{m/s}^2} H^2 \approx 10^{-7}\,\text{m}^{-1} \cdot H^2.$$

Or, if H is in units of km and $\zeta - N$ in units of m:

$$\zeta - N \approx 0.1\,\text{m/km}^2 \cdot H^2.$$

7.5 Difference between orthometric and normal heights

The geoid is the level from which orthometric heights are measured. Therefore, we may write

$$h = H + N,$$

in which h is the height above the reference ellipsoid, and H is the orthometric height.

We may also bring back to memory equation 7.5:

$$h = H^* + \zeta\,,$$

in which ζ is the height anomaly, and H^* is the normal height.

We obtain simply

$$H - H^* = \zeta - N = -\frac{\Delta g_B H}{\gamma}, \tag{7.12}$$

using equation 7.11.

7.6 Calculating orthometric heights precisely

Orthometric heights are a traditional way of expressing height "above sea level". Orthometric heights are heights above a real geoid — a level surface inside the Earth that is, in the mean, located at the same level as the mean sea level.

We may write

$$W = W_0 - \int_0^H g(z)\,dz,$$

in which g is the true gravity inside the topographic masses. From this we obtain

$$H = \frac{C}{\overline{g}} = \frac{-(W - W_0)}{\overline{g}},$$

in which the mean gravity along the plumb line is

$$\overline{g} = \frac{1}{H}\int_0^H g(z)\,dz.$$

The method is recursive: H appears on both the left and right sides. This is not a problem: both H and $\overline{g}$ are obtained iteratively. Convergence is fast.

In practice one calculates orthometric height using an approximate formula. In Finland, *Helmert orthometric heights* have long been used, for which gravity measured on the Earth's surface, $g(H)$, is extrapolated downwards by using the estimated vertical gravity gradient interior to the rock. It is assumed that its standard value outside the rock, the value $-0.3084\,\mathrm{mGal/m}$ (the free-air gradient), changes to a value that is $0.2238\,\mathrm{mGal/m}$ greater (twice the standard-density $2670\,\mathrm{kg/m^3}$ Bouguer-plate effect): the end result is the total inside-rock gravity gradient, $-0.0846\,\mathrm{mGal/m}$. painovoiman ilmagradientti

This is called the *Prey*[5] *reduction*. As the end result we obtain the following equations (the coefficient is *half* the gravity gradient, so the mean gravity along the plumb line is the same as gravity at the midpoint of the plumb line): [5]

$$\overline{g} = g(H) - 0.0846\,\mathrm{mGal/m}\,\left(-\tfrac{1}{2}H\right) = g(H) + 0.0423\,\mathrm{mGal/m}\cdot H,$$

thus

$$H = \frac{C}{\overline{g}} = \frac{C}{g(H) + 0.0423\,\mathrm{mGal/m}\cdot H}, \tag{7.13}$$

in which C is the geopotential number (potential difference with mean sea level) and $g(H)$ is gravity at the Earth's surface. See also Heiskanen and Moritz (1967) pages 163–167. The term $0.0423\,\mathrm{mGal/m}\cdot H$ is typically *much smaller* than $g(H)$, which is about $9.81\,\mathrm{m/s^2} = 981\,000\,\mathrm{mGal}$! So, an iteration in which the denominator is first calculated using a crude H value, converges very fast.

The use of Helmert heights as an approximation to orthometric heights is imprecise for the following reasons:

[5] Adalbert Prey (1873–1949) was an Austrian astronomer and geodesist and an author of textbooks.

- The assumption that gravity changes linearly along the plumb line. This is not the case, especially not because of the effect of the surrounding terrain. In the precise computation of orthometric heights, one should use a sufficient number of support points along the plumb line for computing this effect.
- The assumption that the free-air vertical gravity gradient is everywhere the same, $-0.3084\,{}^{\mathrm{mGal}}/_{\mathrm{m}}$. The real gradient can easily vary by $\pm 10\,\%$ around this value.
- The assumption that the rock density $\rho = 2670\,{}^{\mathrm{kg}}/_{\mathrm{m}^3}$. The true density value may easily vary by $\pm 10\,\%$ or more around this assumed value.

The first approximation, neglecting the terrain effect, can be corrected by using Niethammer's[6] method, see Heiskanen and Moritz (1967) page 167. It requires that, in geoid computation, too, the terrain is correspondingly taken into account.

The third approximation, the density, can be removed as a problem by conventionally agreeing to also use a standard density $\rho = 2670\,{}^{\mathrm{kg}}/_{\mathrm{m}^3}$ in the corresponding geoid determination. The surface thus obtained is not any more a true geoid then, but a "fake geoid", for which no suitable name comes to mind.

The second approximation could be eliminated by using the *true* free-air gravity gradient instead of a standard value. To compute the gradient, the integral equation presented in section 8.6 may be used.

The precise calculation of orthometric heights is thus laborious: just as laborious as the precise determination of the geoid, and for the same reasons. Fortunately in non-mountainous countries, Helmert heights are good enough. In Finland they were even computed using for the ρ values "true" crustal densities according to a geological map (Kääriäinen, 1966, page 32).

[6] Theodor Niethammer (1876–1947) was a Swiss astronomer and geodesist who created the gravimetric base network of Switzerland.

7.7 Calculating normal heights precisely

For this we use equation 7.6:

$$H^* = \frac{C}{\overline{\gamma}} = -\frac{W - W_0}{\overline{\gamma}}, \tag{7.6}$$

where the average value of normal gravity along the plumb line is

$$\overline{\gamma} = \overline{\gamma_{0H}} = \frac{1}{H^*} \int_0^{H^*} \gamma(z)\, dz.$$

Because normal gravity is in good approximation a linear function of z, we may write

$$\overline{\gamma} = \gamma_0 + \tfrac{1}{2} H^* \frac{\partial \gamma}{\partial z},$$

in which $\frac{\partial}{\partial z}\gamma = -0.3084\,\mathrm{mGal/m}$ and $\gamma_0(\varphi) \overset{\text{def}}{=} \gamma(\varphi, 0)$ is normal gravity computed at height zero. We obtain

$$\overline{\gamma} = \gamma_0 - 0.1542\,\mathrm{mGal/m} \cdot H^*.$$

The solution is again obtained iteratively:

$$H^* = \frac{C}{\overline{\gamma}} = \frac{C}{\gamma_0 - 0.1542\,\mathrm{mGal/m} \cdot H^*} \tag{7.14}$$

in which $\gamma_0(\varphi)$ can be calculated exactly when local latitude φ is known. H^* appears on both sides of the equation, but the iterative solution converges fast due to the first term of the denominator γ_0, some $9.81\,\mathrm{m/s^2} = 981\,000\,\mathrm{mGal}$, being a lot larger than $0.1542\,\mathrm{mGal/m} \cdot H^*$.

Calculation of normal heights, unlike calculation of orthometric heights, is not sensitive to Earth crustal density hypotheses. It depends, however, on the choice of normal gravity field, in other words the reference ellipsoid.

7.8 Calculation example for heights

At point P the potential difference with the sea level is $C = 5000\,\mathrm{m^2/s^2}$. Local gravity is $g = 9.820\,000\,\mathrm{m/s^2}$.

Normal gravity calculated at level zero under point P equals $\gamma_0 = 9.821\,500\,\mathrm{m/s^2}$.

Questions

1. Calculate the orthometric height of point P.
2. ilma-anomalia Calculate the free-air gravity anomaly Δg_{FA} of point P.
3. Calculate the Bouguer anomaly (without terrain correction) Δg_B of point P.
4. Calculate the normal height of point P.
5. If the geoid height at point P is $N = 25.000\,\mathrm{m}$, how much is the *height anomaly* ("quasi-geoid height") ζ?

Answers

1. First attempt:

$$H^{(0)} = \frac{C}{g} = \frac{5000}{9.82}\,\mathrm{m} = 519.165\,\mathrm{m}.$$

Second attempt (equation 7.13):

$$H^{(1)} = \frac{5000\,\mathrm{m^2/s^2}}{9.820\,000\,\mathrm{m/s^2} + 0.0423 \cdot 10^{-5}\,\mathrm{s^{-2}} \cdot 519.165\,\mathrm{m}} = \\ = 509.154\,\mathrm{m}.$$

After that, the millimetres no longer change.

2. The free-air anomaly is

$$\Delta g_{FA} = 9.820\,000\,\mathrm{m/s^2} - \\ - \left(9.821\,500 - 0.3084 \cdot 10^{-5} \cdot 509.154\right)\,\mathrm{m/s^2} = \\ = 7.023\,\mathrm{mGal}.$$

3. The Bouguer anomaly is (equation 6.2):

$$\Delta g_B = \Delta g_{FA} - 0.1119\,\mathrm{mGal/m} \cdot H = -49.951\,\mathrm{mGal}.$$

4. The first attempt is again

$$H^{*(0)} = \frac{C}{\gamma_0} = 509.087\,\mathrm{m}.$$

The second, equation 7.14:

$$H^{*(1)} = \frac{5000\,\mathrm{m^2/s^2}}{9.821\,500\,\mathrm{m/s^2} - 0.1542 \cdot 10^{-5}\,\mathrm{s^{-2}} \cdot 509.087\,\mathrm{m}} = \\ = 509.128\,\mathrm{m},$$

also final on the millimetre level.

5. The difference equation 7.12 yields

$$\zeta - N = -\frac{\Delta g_B H}{\gamma} = 0.026\,\mathrm{m}.$$

Also (check) $H - H^* = 0.026\,\mathrm{m}$. So

$$\zeta = N + 0.026\,\mathrm{m} = 25.026\,\mathrm{m}.$$

7.9 Orthometric and normal corrections

In practical orthometric height calculations, one often starts by adding together the height differences ΔH measured by levelling ("staff-reading differences") between points A and B as a *tentative* or crude height difference

$$\sum_{\substack{i=A \\ \text{levelling line}}}^{B-1} \Delta H_{i,i+1} \overset{\text{def}}{=} \sum_{A}^{B} \Delta H,$$

after which the non-exactness of this method is accounted for by applying the "orthometric correction" (OC):

$$H_B = H_A + \sum_{A}^{B} \Delta H + OC_{AB}.$$

The fact that the difference in orthometric heights between two points A and B is not equal to the sum of the levelled height differences is due to gravity not being the same everywhere.

With C_A, C_B and ΔC the geopotential numbers at A and B, and the geopotential differences along the levelling line, it holds that $C_B - C_A -$

$\sum_A^B \Delta C = 0$ because of the conservative nature of the geopotential. Dividing by a constant γ_0 yields

$$\frac{C_B}{\gamma_0} - \frac{C_A}{\gamma_0} - \sum_A^B \frac{\Delta C}{\gamma_0} = 0.$$

On the other hand

$$OC_{AB} = H_B - H_A - \sum_A^B \Delta H = \frac{C_B}{\overline{g}_B} - \frac{C_A}{\overline{g}_A} - \sum_A^B \frac{\Delta C}{g},$$

with $\overline{g}_A$ and $\overline{g}_B$ average gravity values along the plumb lines of A and B and g gravity along the levelling line. This expression compares $\sum_A^B \Delta H$, the naively calculated sum of levelled height differences, with the difference between the orthometric heights of the end points A and B, calculated according to the definition.

Subtraction yields

$$OC_{AB} - 0 = \left(\frac{C_B}{\overline{g}_B} - \frac{C_B}{\gamma_0}\right) - \left(\frac{C_A}{\overline{g}_A} - \frac{C_A}{\gamma_0}\right) - \sum_A^B \left(\frac{\Delta C}{g} - \frac{\Delta C}{\gamma_0}\right),$$

in which

$$\begin{aligned}
\frac{C_B}{\overline{g}_B} - \frac{C_B}{\gamma_0} &= \left(\frac{\gamma_0 - \overline{g}_B}{\gamma_0}\right) \frac{C_B}{\overline{g}_B} = \left(\frac{\gamma_0 - \overline{g}_B}{\gamma_0}\right) H_B, \\
\frac{C_A}{\overline{g}_A} - \frac{C_A}{\gamma_0} &= \left(\frac{\gamma_0 - \overline{g}_A}{\gamma_0}\right) H_A, \\
\frac{\Delta C}{g} - \frac{\Delta C}{\gamma_0} &= \left(\frac{\gamma_0 - g}{\gamma_0}\right) \Delta H,
\end{aligned}$$

yielding the *orthometric correction*

$$OC_{AB} = \sum_A^B \left(\frac{g - \gamma_0}{\gamma_0}\right) \Delta H + \left(\frac{\overline{g}_A - \gamma_0}{\gamma_0}\right) H_A - \left(\frac{\overline{g}_B - \gamma_0}{\gamma_0}\right) H_B, \quad (7.15)$$

which is identical to Heiskanen and Moritz's (1967) equation 4-33.

The choice of the constant γ_0 is arbitrary. It is wise to choose it close to the average gravity in the area of the levelling line AB, so as to keep the numerics small.

Similarly we may also calculate the *normal correction* (NC) in calculating normal heights. Start from the equation

$$NC_{AB} = H^*_B - H^*_A - \sum_A^B \Delta H = \frac{C_B}{\overline{\gamma}_B} - \frac{C_A}{\overline{\gamma}_A} - \sum_A^B \frac{\Delta C}{g}, \tag{7.16}$$

from which, as above, follows by subtraction:

$$NC_{AB} = \sum_A^B \left(\frac{g - \gamma_0}{\gamma_0}\right) \Delta H + \left(\frac{\overline{\gamma}_A - \gamma_0}{\gamma_0}\right) H^*_A - \left(\frac{\overline{\gamma}_B - \gamma_0}{\gamma_0}\right) H^*_B. \tag{7.17}$$

The identical first term in both equation 7.15 and equation 7.17 can be traced back to the term

$$\sum_A^B \frac{\Delta C}{g} = \sum_A^B \Delta H,$$

the naive summation of height differences ΔH in the case of both orthometric and normal correction, which is the generic basis of the concept of both corrections.

Equation 7.16 yields

$$H^*_B = H^*_A + \sum_A^B \Delta H + NC_{AB}.$$

What changes between the orthometric and normal corrections is the definition of heights: H^* instead of H, requiring division by the average of normal gravity along the plumb line $\overline{\gamma}$, not by that of true gravity $\overline{g}$.

Both the orthometric correction 7.15 and the normal correction 7.17 are calculated one benchmark interval at a time: one must know, in addition to the levelled height difference ΔH, local gravity g along the levelling line. Furthermore one must know $g(H)$ or $\gamma(0)$ at both end points in order to calculate mean gravity $\overline{g}$ or $\overline{\gamma}$ along the plumb lines of those end points. All this goes well with the equations given above. Remember that gravity g along the levelling line is needed also if one wants to reduce the individual levelled height differences ΔH to geopotential number differences ΔC. This reduction is part of the computation of both the orthometric and the normal correction.

7.10 A vision for the future: relativistic levelling

According to general relativity, the deeper a clock is inside the potential well of masses, the slower it ticks. This is most easily seen by looking at the Schwarzschild[7] metric for a spherically symmetric field:

7 metriikka

$$\begin{aligned} c^2\,d\tau^2 &= \\ &= \left(1-\frac{2GM}{c^2 r}\right)c^2\,dt^2 - \left(1-\frac{2GM}{c^2 r}\right)^{-1} dr^2 - r^2\left(d\phi^2+\cos^2\phi\,d\lambda^2\right) = \\ &= \left(1-\frac{2W}{c^2}\right)c^2\,dt^2 - \left(1-\frac{2W}{c^2}\right)^{-1} dr^2 - r^2\left(d\phi^2+\cos^2\phi\,d\lambda^2\right), \end{aligned}$$

in spherical co-ordinates plus time (ϕ, λ, r, t). Here we see how the rate of proper time τ is slowed down compared to stationary co-ordinate time t (time at infinity $r \to \infty$), when the geopotential W increases closer to the mass. The slowing-down ratio is

ominaisaika

$$\frac{\partial\tau}{\partial t} = \sqrt{1-\frac{2W}{c^2}} \approx 1-\frac{W}{c^2}.$$

Now c^2, the speed of light squared, is, in the units of daily life, a huge number: $10^{17}\,m^2/s^2$. This means that measuring a potential difference of $1\,m^2/s^2$ — corresponding to a height difference of 10 cm — using this method, requires a precision of $1:10^{17}$. More traditional, microwave-based atomic clocks can do precisions of 10^{-12}–10^{-14} (Vermeer, 1983a). With the new optical clocks, the objective should be achievable and relativistic levelling may become a reality.

The clock works in this way: extreme cooling produces a so-called Bose–Einstein condensate of atoms, which is trapped inside an optical lattice formed by six laser beams, an electromagnetic pattern of standing waves. The clock oscillation is on a different frequency. A Bose–Einstein

valohila

[7] Karl Schwarzschild (1873–1916) was a German physicist who was the first to derive, in 1915 while serving on the Russian front, a closed spherically symmetric, non-rotating solution to the field equation of Albert Einstein's general theory of relativity, the *Schwarzschild metric*.

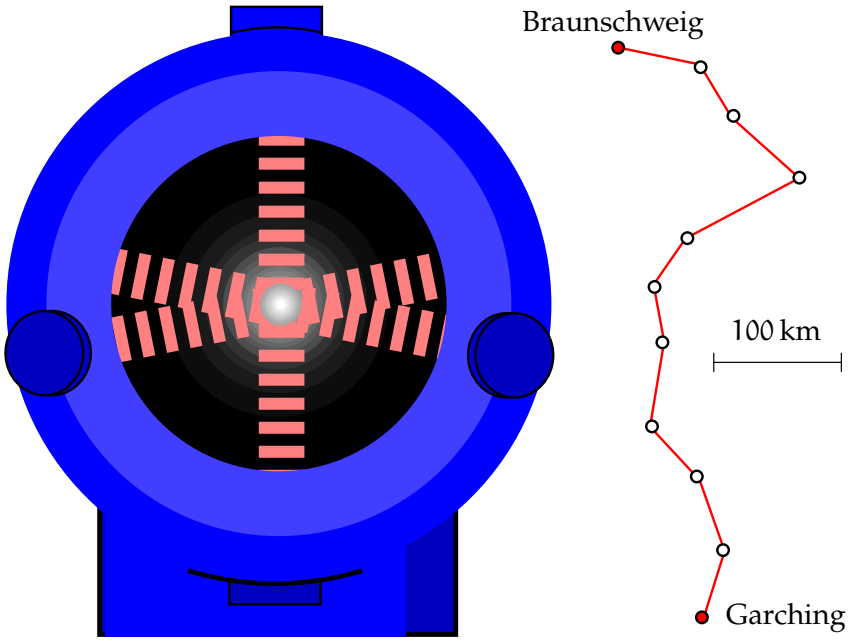


FIGURE 7.8. An optical lattice clock: the ultra-precise atomic clock of the future operates at optical wavelengths. To the right, the trajectory of the Predehl et al. (2012) experiment.

condensate has the property that all atoms are in precisely the same quantum state — like the photons in an operating laser: their matter waves are *coherent*. In a way, all the atoms together act as one virtual atom. The condensate may consist of millions of atoms.

Unfortunately it is not enough that just one laboratory measures time to extreme precision. One has also to be able to *compare* the ticking rates of different clocks over geographical distances. For this, a solution has also been found: existing fibreoptic cables already in global use for Internet and telephony are useable for this with small modifications. The modifications concern the amplifiers in the cables at distances of some 100 km, which must be replaced by modified ones (Predehl et al., 2012). In this way, both the traditional precise levelling networks and the height systems based on GNSS technology and geoid determination may be replaced by this hi-tech (and hi-science!) solution.

Self-test questions

1. Why are heights calculated directly from levelled height differences not good enough as a height *system*?
2. What is a geopotential number?
3. What are orthometric heights?
4. What are normal heights?
5. What is the classical definition of the geoid?
6. What is a height anomaly?
7. What is the quasi-geoid?
8. Why might water sometimes flow in the "wrong" direction, to a greater height?
9. What is the telluroid?
10. What are the orthometric correction and the normal correction?

Exercise 7−1: Calculating orthometric heights

The potential difference with sea level at point P, $-(W - W_0)$, equals $1000\,\mathrm{m^2/s^2}$. Gravity at the point is $g_P = 9.820\,000\,\mathrm{m/s^2}$. Calculate the orthometric height of the point. Aim for millimetre precision.

Exercise 7−2: Calculating normal heights

At point P, the potential difference with sea level is

$$-(W - W_0) = 5000\,\mathrm{m^2/s^2}.$$

Below the point at sea level, normal gravity is $\gamma_0 = 9.821\,500\,\mathrm{m/s^2}$. Calculate the normal height of the point.

Exercise 7–3: Difference between orthometric and normal height

At point P, the Bouguer anomaly is $\Delta g_B = -120\,\text{mGal}$. The orthometric height of the point is 1150 m.

1. Calculate the normal height of point P.
2. If the geoid height in point P is $N = 21.75\,\text{m}$, calculate the height anomaly ζ of the point.

The Stokes equation and other integral equations

8

8.1 The Stokes equation and the Stokes integral kernel

Assume a spherical Earth. By suitably combining the equations in section 5.3, one obtains at sea level

$$T = \sum_{n=2}^{\infty} T_n = R \sum_{n=2}^{\infty} \frac{\Delta g_n}{n-1},$$

with $T_n = T_n(\phi, \lambda)$ the degree constituents of the disturbing potential field $T = T(\phi, \lambda)$, and $\Delta g_n = \Delta g_n(\phi, \lambda)$ those of the gravity anomaly field $\Delta g = \Delta g(\phi, \lambda)$. The summation starts from $n = 2$; for the degree numbers $n = 0, 1$, the Δg_n are assumed to vanish, as $\Delta g_0 \neq 0$ would mean a different total mass for the normal field than for the Earth, and $\Delta g_1 \neq 0$ an offset of the co-ordinate origin from the Earth's centre of mass, see section 3.4.

häiriö-potentiaali

This is now the Stokes equation's *spectral form*.

Substituting into this degree constituent equation 3.9, one obtains the integral equation

asteosuus-yhtälö

$$T = \frac{R}{4\pi} \sum_{n=2}^{\infty} \frac{2n+1}{n-1} \iint_{\sigma} \Delta g(\phi', \lambda')\, P_n(\cos\psi)\, d\sigma' =$$

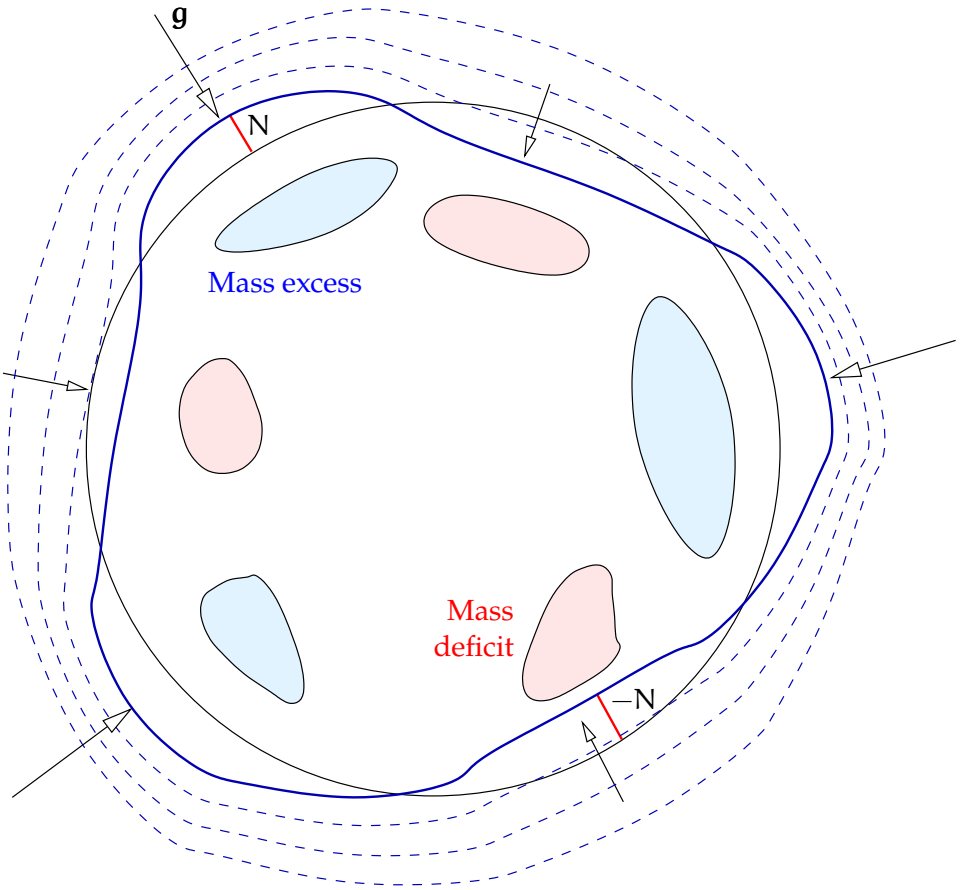


FIGURE 8.1. The principle of gravimetric geoid determination.

$$= \frac{R}{4\pi} \iint_\sigma \left(\sum_{n=2}^{\infty} \frac{2n+1}{n-1} P_n(\cos\psi) \right) \Delta g(\phi', \lambda')\, d\sigma' = \\ = \frac{R}{4\pi} \iint_\sigma S(\psi)\, \Delta g(\phi', \lambda')\, d\sigma', \quad (8.1)$$

in which

$$S(\psi) = \sum_{n=2}^{\infty} \frac{2n+1}{n-1} P_n(\cos\psi),$$

the Stokes kernel function. The angle ψ is the geocentric angular distance between the evaluation point and moving observation point, see figure 8.2. The equation above allows the calculation, from global gravimetric data and for every point on the surface of the Earth sphere,

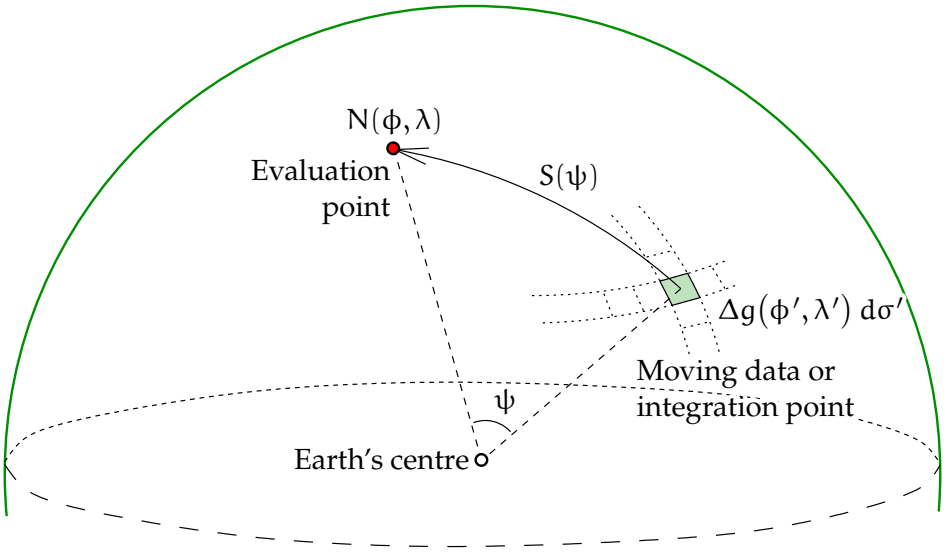


FIGURE 8.2. The geometry of integrating the Stokes equation.

of the disturbing potential T, and from that, the *geoid height* N using Bruns equation 5.2, $N = T/\gamma$. The result is

$$N(\phi,\lambda) = \frac{T(\phi,\lambda)}{\gamma} = \frac{R}{4\pi\gamma}\iint_\sigma S(\psi)\,\Delta g(\phi',\lambda')\,d\sigma', \qquad (8.2)$$

in which (ϕ,λ) and (ϕ',λ') are the evaluation point and the moving point ("data point"), respectively, and the angular distance between them is ψ. Equation 8.2 is the classical Stokes integral equation of gravimetric geoid determination.

The above illustrates the correspondence between integral equations and spectral expansions. There are other examples of this, like the spectral representation of the function $1/\ell$, equation 8.7, Heiskanen and Moritz's (1967) equation 1-81. Of course $1/\ell$ is also the kernel function of an integral equation, equation 1.28, the one yielding the potential V if the single-layer mass density κ is given.

A version of the Stokes equation for the *exterior space* also exists. We gave it earlier, equation 5.9. The spectral form of its kernel function is

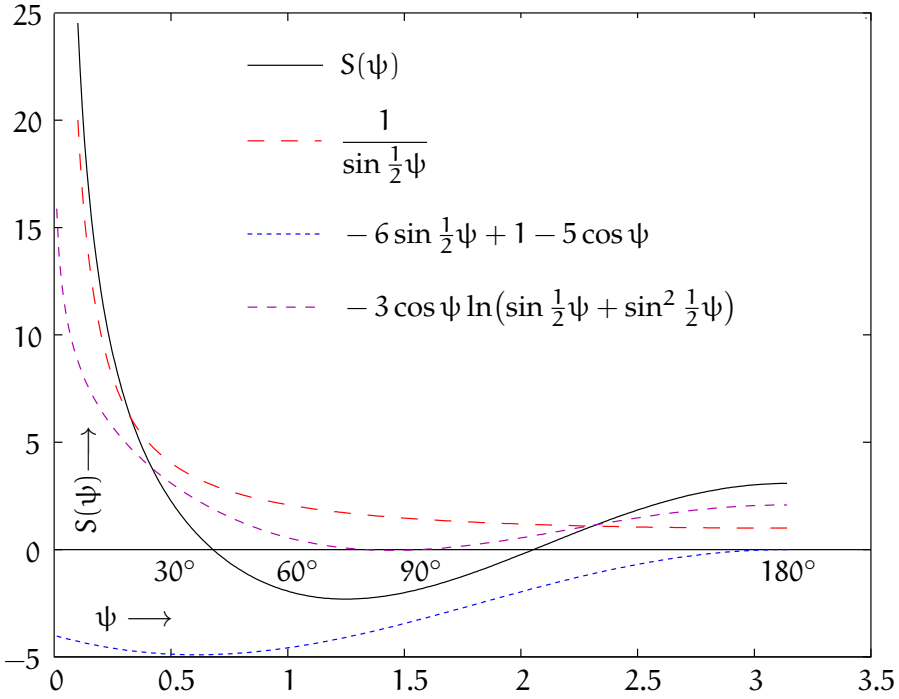


FIGURE 8.3. The Stokes kernel function $S(\psi)$. The argument ψ is in radians $[0, \pi)$. Also plotted are the three parts of analytical expression 8.3 with their different asymptotic behaviours.

equation 5.10:

$$S(\psi, r, R) = \sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+1} \frac{2n+1}{n-1} P_n(\cos\psi). \qquad (5.10)$$

The Stokes kernel function on the Earth's surface is depicted in figure 8.3, in which the angle ψ is in radians (1 rad $\approx 57^\circ.29578\ldots$). The curve was calculated using the following closed expression (Heiskanen and Moritz, 1967, section 2-16, equation 2-164):

$$S(\psi) = \frac{1}{\sin\frac{1}{2}\psi} - 6\sin\tfrac{1}{2}\psi + 1 - 5\cos\psi - \\ - 3\cos\psi \ln\left(\sin\tfrac{1}{2}\psi + \sin^2\tfrac{1}{2}\psi\right). \qquad (8.3)$$

This closed expression helps us to understand better how the function behaves close to the origin $\psi = 0$: the first term, $1/\sin\frac{1}{2}\psi$, goes to

infinity when $\psi \to 0$. The next three terms, $-6\sin\frac{1}{2}\psi + 1 - 5\cos\psi$, are all bounded on the whole interval $[0, \pi)$ and the value for $\psi = 0$ is -4. The last, complicated term $-3\cos\psi\ln\left(\sin\frac{1}{2}\psi + \sin^2\frac{1}{2}\psi\right)$ also goes to infinity — *positive* infinity! — for $\psi \to 0$, but much more slowly because of the logarithm.

8.2 Example: The Stokes equation in polar co-ordinates

In section 2.3 we derived a general solution to the Laplace equation in two dimensions in polar co-ordinates. Below we develop a "toy" computation framework for gravimetric geoid determination in two dimensions, which allows us to do simple numerical simulations in order to get a feel for the behaviour of these things.

Firstly we derive the disturbing potential, gravity anomaly, and Stokes integral kernel for this solution, equation 2.3, assuming a normal potential $U(r) = a_0 + b_0 \ln r$.

- Disturbing potential:

$$T(\alpha, r) = V^{\text{ext}}(\alpha, r) - (a_0 + b_0 \ln r) = \\ = \sum_{k=1}^{\infty} r^{-k} \left(a_k \cos k\alpha + b_k \sin k\alpha\right).$$

- Normal gravity:

$$\gamma(r) = -\frac{\partial U}{\partial r} = -\frac{b_0}{r}.$$

- Normal gravity gradient:

$$\frac{\partial \gamma}{\partial r} = -\frac{\partial^2 U}{\partial r^2} = \frac{b_0}{r^2}.$$

- Gravity anomaly, equation 5.5:

$$\Delta g(\alpha, r) = -\frac{\partial T}{\partial r} + \frac{1}{\gamma}\frac{\partial \gamma}{\partial r} T =$$

$$= \sum_{k=1}^{\infty} \frac{k}{r} r^{-k} \left(a_k \cos k\alpha + b_k \sin k\alpha \right) -$$

$$- \frac{1}{r} \sum_{k=1}^{\infty} r^{-k} \left(a_k \cos k\alpha + b_k \sin k\alpha \right) =$$

$$= \sum_{k=2}^{\infty} \frac{k-1}{r} r^{-k} \left(a_k \cos k\alpha + b_k \sin k\alpha \right).$$

We see that, if we write

$$T(\alpha, r) = \sum_{k=1}^{\infty} \left(\frac{R}{r} \right)^k T_k(\alpha), \quad T_k(\alpha) \stackrel{\text{def}}{=} R^{-k} \left(a_k \cos k\alpha + b_k \sin k\alpha \right),$$

it follows that

$$\Delta g(\alpha, r) = \sum_{k=2}^{\infty} \left(\frac{R}{r} \right)^{k+1} \Delta g_k(\alpha),$$

$$\Delta g_k(\alpha) \stackrel{\text{def}}{=} (k-1) R^{-(k+1)} \left(a_k \cos k\alpha + b_k \sin k\alpha \right),$$

and, like in the case of spherical co-ordinates,

$$\Delta g_k(\alpha) = \frac{k-1}{R} T_k(\alpha). \tag{8.4}$$

kantafunktio According to Fourier theory, the basis functions $\cos k\alpha$ and $\sin k\alpha$ are *orthonormal* on the circle $r = R$ when choosing the following integral as the scalar product:

$$\frac{1}{\pi} \int_0^{2\pi} \cos k\alpha \cos m\alpha \, d\alpha = \frac{1}{\pi} \int_0^{2\pi} \sin k\alpha \sin m\alpha \, d\alpha = \begin{cases} 0 & \text{if } k \neq m, \\ 1 & \text{if } k = m, \end{cases}$$

$$\frac{1}{\pi} \int_0^{2\pi} \cos k\alpha \sin m\alpha \, d\alpha = 0 \quad \text{always.}$$

This means that we may expand

$$\Delta g(\alpha, R) = \sum_{k=2}^{\infty} \Delta g_k(\alpha)$$

into its Fourier terms as follows:

$$\Delta g_k(\alpha) \stackrel{\text{def}}{=} (k-1)\,R^{-(k+1)}\,(a_k \cos k\alpha + b_k \sin k\alpha) = \\ = \overbrace{(k-1)\,R^{-(k+1)} a_k}^{A_k} \cos k\alpha + \overbrace{(k-1)\,R^{-(k+1)} b_k}^{B_k} \sin k\alpha.$$

This yields the following Fourier coefficients:

$$\left\{ \begin{array}{c} A_k \\ B_k \end{array} \right\} = (k-1)\,R^{-(k+1)} \left\{ \begin{array}{c} a_k \\ b_k \end{array} \right\}, \qquad k = 2, 3, \cdots$$

and on the circle $r = R$ the expansion is

$$\Delta g(\alpha, R) = \sum_{k=2}^{\infty} \Delta g_k(\alpha) = \sum_{k=2}^{\infty} (A_k \cos k\alpha + B_k \sin k\alpha)\,.$$

The substitutions

$$\left\{ \begin{array}{c} a_k \\ b_k \end{array} \right\} = \frac{R^{k+1}}{k-1} \left\{ \begin{array}{c} A_k \\ B_k \end{array} \right\}$$

yield

$$T(\alpha, R) = \sum_{k=2}^{\infty} T_k(\alpha) = \sum_{k=2}^{\infty} R^{-k}\,(a_k \cos k\alpha + b_k \sin k\alpha) = \\ = \sum_{k=2}^{\infty} R^{-k} \left(\frac{R^{k+1}}{k-1} A_k \cos k\alpha + \frac{R^{k+1}}{k-1} B_k \sin k\alpha \right) = \\ = \sum_{k=2}^{\infty} \frac{R}{k-1} (A_k \cos k\alpha + B_k \sin k\alpha)\,.$$

Using the equations for the Fourier coefficients,

$$\left\{ \begin{array}{c} A_k \\ B_k \end{array} \right\} = \frac{1}{\pi} \int_0^{2\pi} \Delta g(\alpha, R) \left\{ \begin{array}{c} \cos k\alpha \\ \sin k\alpha \end{array} \right\} d\alpha,$$

and the cosine difference equation (Wolfram Demonstrations, Difference formula for cosine), we obtain

$$T(\alpha, R) = \frac{1}{\pi} \cdot \cdot \sum_{k=2}^{\infty} \frac{R}{k-1} \left(\cos k\alpha \int_0^{2\pi} \Delta g(\alpha', R) \cos k\alpha' \, d\alpha' + \sin k\alpha \int_0^{2\pi} \Delta g(\alpha', R) \sin k\alpha' \, d\alpha' \right) = = \frac{1}{\pi} \sum_{k=2}^{\infty} \frac{R}{k-1} \int_0^{2\pi} \Delta g(\alpha', R) \cdot \cos\big(k\,(\alpha - \alpha')\big) \, d\alpha'.$$

Define the Stokes kernel for this two-dimensional situation:

$$N(\alpha) = \frac{T(\alpha, R)}{\gamma} = \frac{R}{\pi\gamma} \int_0^{2\pi} \Delta g(\alpha', R)\, S\big(\alpha - \alpha'\big) \, d\alpha',$$

$$\text{with} \quad S\big(\alpha - \alpha'\big) \stackrel{\text{def}}{=} \sum_{k=2}^{\infty} \frac{\cos\big(k\,(\alpha - \alpha')\big)}{k-1}.$$

For small values of $\alpha - \alpha'$ we may approximate (Wolfram Functions, $\sum_{k=1}^{\infty} \frac{\cos kx}{k}$):

$$S(\alpha - \alpha') = \sum_{k'=1}^{\infty} \frac{\cos\big((k'+1)\,(\alpha - \alpha')\big)}{k'} \approx \sum_{k'=1}^{\infty} \frac{\cos\big(k'\,(\alpha - \alpha')\big)}{k'} = = \frac{1}{2} \ln\left(\frac{1}{2\,(1 - \cos(\alpha - \alpha'))} \right) \approx -\ln(\alpha - \alpha').$$

More abstractly, we may also write relationship 8.4 in terms of the discrete Fourier transform and its inverse, as

$$\mathcal{F}\{\Delta g\} = \frac{k-1}{R} \mathcal{F}\{T\} \quad \Longrightarrow \quad T = \mathcal{F}^{-1}\left\{ \frac{R}{k-1} \mathcal{F}\{\Delta g\} \right\}.$$

Here, $\mathcal{F}\{f\}$ represents the Fourier transform of a function $f(\alpha)$ of spatial co-ordinate α on the circle, into a function of the spatial wave number (number of waves around the circle) k.

This formulation has the merit of being able to use any standard FFT software library offering compatible versions of both the forward Fourier transform $\mathcal{F}\{\cdot\}$ and the inverse transform $\mathcal{F}^{-1}\{\cdot\}$.

For more about FFT, see appendix C.

Figure 8.5 shows a simulation result in which a randomly generated set of gravity anomalies on the circle $r = R$ has been used to calculate

TABLEAU 8.1. Stokes equation in two dimensions, octave code.

```
% Stokes equation simulator in two dimensions
R = 6378137;
g = 9.8;
ak(1:180) = 0.0;
bk(1:180) = 0.0;
dg(1:360) = 0.0;
T(1:360) = 0.0;
for i=1:359
% Gauss-Markov
  dg(i+1) = 0.8*dg(i) + 50*(rand()-0.5);
end
dgsum = 0.0;
for i=1:360
  % Enforce circularity
  dg(i) = dg(i) - (dg(360) - dg(1)) * (i/359);
  dgsum = dgsum + dg(i);
end
for i = 1:360
  % Enforce zero expectancy
  dg(i) = dg(i) - dgsum/360;
  for k = 2:180
    ak(k) = ak(k) + dg(i) * cos(k*i*pi/180)/180;
    bk(k) = bk(k) + dg(i) * sin(k*i*pi/180)/180;
  end
end
dg(1:360) = 0.0;
for i=1:360
  for k = 2:180
    T(i) = T(i) + (ak(k)*cos(k*i*pi/180) + bk(k)*sin(k*i*pi/180))*R/(k-1);
    % Without degree one
    dg(i) = dg(i) + ak(k)*cos(k*i*pi/180) + bk(k)*sin(k*i*pi/180);
 end
end
hold on
plot(1:360, dg, 'b') plot(1:360, 0.00001*T/g, 'm')
print -dpdf stokes2D-out.pdf
```

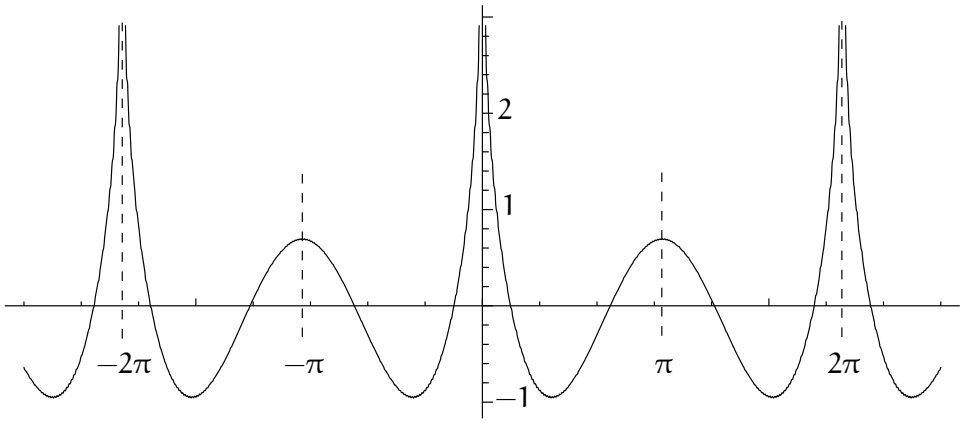


FIGURE 8.4. The Stokes kernel function on the circle $r = R$ in two-dimensional geometry. Note the symmetry and periodicity. Compare with the spherical Stokes kernel, figure 8.3.

geoid undulations on the same circle. Both curves display fairly realistic statistical behaviour. The code used is given in tableau 8.1.

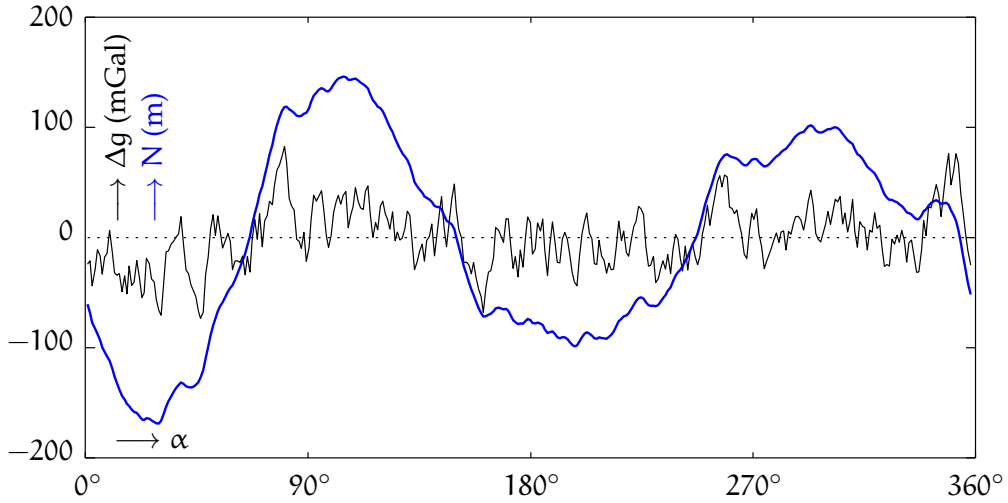


FIGURE 8.5. Simulation of gravity anomalies (Gauss–Markov process) and geoid undulations (blue) in two-dimensional geometry on the circle. Note the spectral behaviour of both.

8.3 Plumb-line deflections and Vening Meinesz equations

By differentiating the Stokes equation with respect to place, we obtain integral equations for the components of the deflection of the plumb line (Heiskanen and Moritz, 1967, equation 2-210′):

luotiviivan poikkeama

$$\left\{ \begin{array}{c} \xi(\phi,\lambda) \\ \eta(\phi,\lambda) \end{array} \right\} = \frac{1}{4\pi\gamma} \iint_\sigma \Delta g(\phi',\lambda') \frac{dS(\psi)}{d\psi} \left\{ \begin{array}{c} \cos\alpha \\ \sin\alpha \end{array} \right\} d\sigma' =$$

$$= \frac{1}{4\pi\gamma} \iint_\sigma \Delta g(\phi',\lambda') \frac{dS(\psi)}{d\psi} \left\{ \begin{array}{c} \cos\alpha \\ \sin\alpha \end{array} \right\} \sin\psi \, d\alpha \, d\psi, \quad (8.5)$$

in which ξ and η are the south-north and west-east direction deflections of the plumb line, and the unit-sphere surface element is $d\sigma' = \sin\psi \, d\alpha \, d\psi$, in which $\sin\psi$ is Jacobi's determinant of the (ψ, α) co-ordinates.

These equations were derived for the first time by the Dutch geophysicist F. A. Vening Meinesz. The angle α is the azimuth or direction angle between the calculation or evaluation point (ϕ, λ) and the moving integration or observation point (ϕ', λ'). These equations are much harder to write in spectral form, as the kernel functions are now also functions of the azimuth direction α; in other words, they are *anisotropic*.

The disturbing potential, the gravity disturbance, and the gravity anomaly are all so-called *isotropic* quantities: they do not depend on the azimuth and therefore, in the spectral representation the transformations between them are functions of harmonic degree n only.

8.4 The Poisson integral equation

Look at figure 8.6. The point Q of the body is located at $\mathbf{R}$, and the observation point P at $\mathbf{r}$. The geocentric angular distance between the two location vectors, in other words the angular distance as seen from the origin, is ψ. The distance between points P and Q is ℓ.

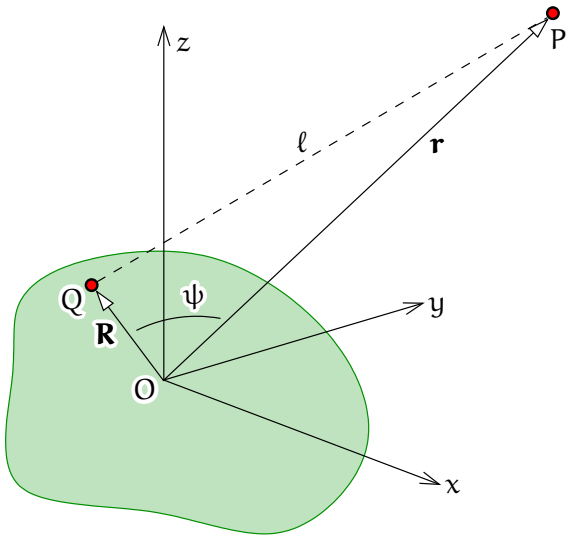


FIGURE 8.6. The geometry of the generating function of the Legendre polynomials.

With the definitions $R \stackrel{\text{def}}{=} \|\mathbf{R}\|$ and $r \stackrel{\text{def}}{=} \|\mathbf{r}\|$, we may write (cosine rule):

$$\ell^2 = r^2 + R^2 - 2rR\cos\psi. \tag{8.6}$$

We may also write the function $1/\ell$ as the following expansion (Heiskanen and Moritz, 1967, equation 1-81):

$$\frac{1}{\ell} = \frac{1}{\sqrt{r^2 + R^2 - 2Rr\cos\psi}} = \frac{1}{R}\sum_{n=0}^{\infty}\left(\frac{R}{r}\right)^{n+1} P_n(\cos\psi), \tag{8.7}$$

in which r and R are the distances of points P and Q from the origin O, usually the centre of the Earth. The function $1/\ell$ is called the *generating function* of the Legendre polynomials.

Differentiating equation 8.7 with respect to r yields

$$-\frac{r - R\cos\psi}{\ell^3} = -\frac{1}{R}\sum_{n=0}^{\infty}\frac{n+1}{r}\left(\frac{R}{r}\right)^{n+1} P_n(\cos\psi).$$

Multiply this by $2r$:

$$-\frac{2r^2 - 2rR\cos\psi}{\ell^3} = -\frac{1}{R}\sum_{n=0}^{\infty}(2n+2)\left(\frac{R}{r}\right)^{n+1} P_n(\cos\psi).$$

Now add together this equation and equation 8.7:

$$\frac{-2r^2 + 2rR\cos\psi + \ell^2}{\ell^3} = -\frac{1}{R}\sum_{n=0}^{\infty}(2n+1)\left(\frac{R}{r}\right)^{n+1} P_n(\cos\psi).$$

Substitute ℓ^2 from equation 8.6:

$$\frac{-2r^2 + 2rR\cos\psi + \ell^2}{\ell^3} = \frac{R^2 - r^2}{\ell^3},$$

and the result is, after multiplying with $-R$,

$$\frac{R\left(r^2 - R^2\right)}{\ell^3} = \sum_{n=0}^{\infty}(2n+1)\left(\frac{R}{r}\right)^{n+1} P_n(\cos\psi). \tag{8.8}$$

Applying degree constituent equation 3.9 to the harmonic potential field V on the spherical Earth's surface, radius R:

$$V_n(\phi,\lambda) = \frac{2n+1}{4\pi}\iint_\sigma V(\phi',\lambda',R)\,P_n(\cos\psi)\,d\sigma',$$

as well as the spectral expansion of the field in space 3.10:

$$V(\phi,\lambda,r) = \sum_{n=0}^{\infty}\left(\frac{R}{r}\right)^{n+1} V_n(\phi,\lambda),$$

we obtain

$$\begin{aligned} V(\phi,\lambda,r) &= \\ &= \frac{1}{4\pi}\sum_{n=0}^{\infty}\left(\frac{R}{r}\right)^{n+1}(2n+1)\iint_\sigma V(\phi',\lambda',R)\,P_n(\cos\psi)\,d\sigma' = \\ &= \frac{1}{4\pi}\iint_\sigma V(\phi',\lambda',R)\left[\sum_{n=0}^{\infty}(2n+1)\left(\frac{R}{r}\right)^{n+1} P_n(\cos\psi)\right] d\sigma' = \\ &= \frac{1}{4\pi}\iint_\sigma \frac{R\left(r^2 - R^2\right)}{\ell^3} V(\phi',\lambda',R)\,d\sigma' \end{aligned}$$

by replacing the expression in square brackets by equation 8.8.

Thus we have obtained the Poisson integral for computing a harmonic field V from values given on the Earth's surface:

$$V_P = \frac{1}{4\pi} \iint_\sigma \frac{R\left(r^2 - R^2\right)}{\ell^3} V_Q \, d\sigma_Q, \tag{8.9}$$

in which ℓ is again the straight-line distance between evaluation point P (where V_P is being computed) and moving data point Q (on the surface of the sphere, V_Q under the integral sign). In this equation we have given the points symbolic names: the co-ordinates of evaluation point P are (ϕ, λ, r), the co-ordinates of data point Q are (ϕ', λ', R).

Still a third way to write the same equation, useful when the harmonic function or field V is not actually defined between the topographic Earth's surface and sea level, is

$$V = \frac{1}{4\pi} \iint_\sigma \frac{R\left(r^2 - R^2\right)}{\ell^3} V^* \, d\sigma.$$

Here, V^* denotes the value of a *harmonically downwards continued* function V — downwards continued into the topography, all the way down to sea level, or, in spherical approximation, to the surface of the sphere $r = R$. This is a function that above the topography is identical to V, is harmonic, and exists also between the topography and sea level. The question of the existence of such a function has been a classical theoretical nut to crack. . . .

reuna-arvo-tehtävä

Equation 8.9 solves for this special case the so-called *Dirichlet boundary-value problem*, finding a harmonic function in an area of space when the value of the function on the boundary of the area has been given.

8.5 Gravity anomalies in the exterior space

The equation derived in section 8.4, equation 8.9, applies for an arbitrary *harmonic* field V, meaning any field for which $\Delta V = 0$. The equation applies conveniently to the expression $r\,\Delta g$, the gravity anomaly multiplied by the geocentric radius, which is also a harmonic field. This is how

we can express the gravity anomaly in the external space $\Delta g(\phi, \lambda, r)$ as a function of gravity anomalies $\Delta g(\phi', \lambda', R)$ on a reference sphere of radius R. The function $r\Delta g$ is harmonic, because according to equation 5.7 vertauspallo

$$\Delta g = \frac{1}{r} \sum_{n=2}^{\infty} (n-1) \left(\frac{R}{r}\right)^{n+1} T_n,$$

so

$$r\Delta g = \sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+1} (n-1)\, T_n = \sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+1} T'_n,$$

in which $T'_n(\phi, \lambda) = (n-1)\, T_n(\phi, \lambda)$ is a perfectly legal surface spherical harmonic just like $T_n(\phi, \lambda)$ itself: the dependence on the radius r, the factor $\left(R/r\right)^{n+1}$, is the same as for the (harmonic) potential. So, Poisson's integral equation 8.9 applies to function $r\Delta g$: pinta-pallofunktio

$$\left[r\Delta g(\phi, \lambda, r)\right] = \frac{1}{4\pi} \iint_{\sigma} \frac{R\left(r^2 - R^2\right)}{\ell^3} \left[R\Delta g\left(\phi', \lambda', R\right)\right] d\sigma'$$

or

$$\Delta g(\phi, \lambda, r) = \frac{1}{4\pi} \iint_{\sigma} \frac{R}{r} \frac{R\left(r^2 - R^2\right)}{\ell^3} \Delta g\left(\phi', \lambda', R\right) d\sigma'. \tag{8.10}$$

An alternative notation is

$$\Delta g = \frac{1}{4\pi} \iint_{\sigma} \frac{R}{r} \frac{R\left(r^2 - R^2\right)}{\ell^3} \Delta g^* \, d\sigma,$$

in which Δg^* denotes the gravity anomaly at sea level, again calculated *by harmonic downwards continuation* of the exterior field, in this case the expression $r\Delta g$.

From equation 8.10 we may lift the closed form of the kernel, which is dimensionless:

$$K(\ell, r, R) = \frac{R}{r} \frac{R\left(r^2 - R^2\right)}{\ell^3},$$

with which

$$\Delta g(\phi, \lambda, r) = \frac{1}{4\pi} \iint_{\sigma} K(r, \psi, R)\, \Delta g\left(\phi', \lambda', R\right) d\sigma'.$$

Using the approximation $r + R \approx 2r$ still yields

$$\Delta g(\phi, \lambda, r) \approx \frac{1}{2\pi} \iint_\sigma R^2 \frac{r - R}{\ell^3} \Delta g(\phi', \lambda', R)\, d\sigma'.$$

Alternatively, we derive the *spectral form*:

$$\Delta g(\phi, \lambda, r) = \\ = \frac{1}{r} \sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+1} (n-1)\, T_n(\phi, \lambda) = \sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+2} \Delta g_n(\phi, \lambda).$$

Degree constituent equation 3.9 gives the functions Δg_n:

$$\Delta g_n(\phi, \lambda) = \frac{2n+1}{4\pi} \iint_\sigma \Delta g(\phi', \lambda', R)\, P_n(\cos\psi)\, d\sigma',$$

with the aid of which

$$\Delta g(\phi, \lambda, r) = \\ = \frac{1}{4\pi} \sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+2} (2n+1) \iint_\sigma \Delta g(\phi', \lambda', R)\, P_n(\cos\psi)\, d\sigma' = \\ = \frac{1}{4\pi} \iint_\sigma \left(\sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+2} (2n+1)\, P_n(\cos\psi) \right) \Delta g(\phi', \lambda', R)\, d\sigma' = \\ = \frac{1}{4\pi} \iint_\sigma K_{\text{mod}}\, \Delta g(\phi', \lambda', R)\, d\sigma',$$

in which

$$K_{\text{mod}}(\psi, r, R) \stackrel{\text{def}}{=} \sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+2} (2n+1)\, P_n(\cos\psi)$$

is the modified spectral version of the Poisson kernel for gravity anomalies. From this kernel, the constituents of degree number 0 and 1 have been removed, see Heiskanen and Moritz (1967) equation 2-159.

Compared to the Stokes kernel, the Poisson kernel drops off fast to zero for growing values of ℓ. In other words, the evaluation of the integral equation may be restricted to a very local area, like a cap of

kalotti

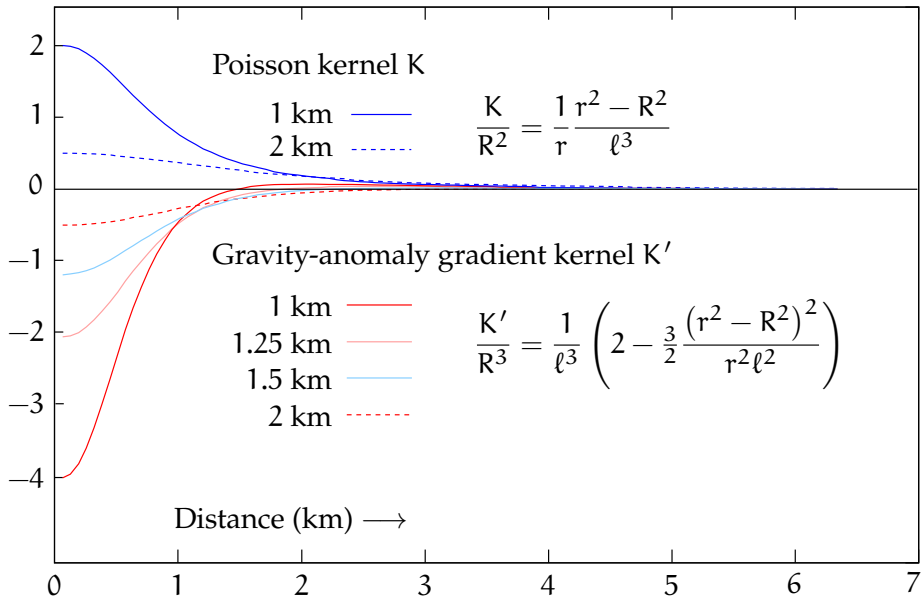


FIGURE 8.7. The Poisson kernel function for gravity anomalies as well as the kernel for the anomalous vertical gravity gradient, both for various height differences $r - R$. These kernels are used when evaluating the surface integral in map co-ordinates (x, y) in kilometres.

radius 1°. See figure 8.7. The main use of Poisson's kernel is the *harmonic continuation*, upwards or downwards, of gravity anomalies measured and computed at various levels, shifting them to the same reference level. vertaustaso

In the limit $r \to R$ (sea level becomes the level of evaluation), this kernel function goes asymptotically to the two-dimensional Dirac δ function. This is inevitable for a kernel that computes gravity anomalies from gravity anomalies.

8.6 The vertical gradient of the gravity anomaly

Differentiate an equation obtained from equations 5.8 and 5.7:

$$\Delta g = \sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+2} \Delta g_n \quad \Longrightarrow \quad \frac{\partial \Delta g}{\partial r} = -\frac{1}{R} \sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+3} (n+2)\, \Delta g_n .$$

This equation is exact in spherical approximation. Its kernel function is well *localised*, in other words, it drops off to zero very fast. For calculation, a small "cap" also suffices here.

Δg_n is expressed, using degree constituent equation 3.9, as an integral over the anomaly field at sea level:

$$\Delta g_n(\phi,\lambda) = \frac{2n+1}{4\pi} \iint_\sigma \Delta g(\phi',\lambda',R)\, P_n(\cos\psi)\, d\sigma',$$

so

$$\frac{\partial \Delta g(\phi,\lambda,r)}{\partial r} = -\frac{1}{4\pi R} \sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+3} (2n+1)\,(n+2) \iint_\sigma \Delta g(\phi',\lambda',R)\, P_n(\cos\psi)\, d\sigma' = \\ = \frac{1}{4\pi R} \iint_\sigma K'(\psi,r,R)\, \Delta g(\phi',\lambda',R)\, d\sigma', \quad (8.11)$$

in which the (dimensionless) kernel function is now

$$K'(\psi,r,R) = -\sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+3} (2n+1)\,(n+2)\, P_n(\cos\psi).$$

Alternatively, derive a closed expression. Start from Poisson equation 8.10 for gravity anomalies, and differentiate with respect to r. See tableau 8.2.

In the result, the last term is small, less than one part in a thousand, compared to the preceding term.

The terms inside the square brackets require their own consideration. In the local zone $\ell \approx r - R$ the terms are similar in magnitude; the second term goes however rapidly to zero for $\ell \gg r - R$. The factor $1/\ell^3$ does so however even more rapidly.

Write

$$\frac{\partial \Delta g(\phi,\lambda,r)}{\partial r} = \frac{R^2}{4\pi} \iint_\sigma \kappa \frac{\Delta g(\phi',\lambda',R)}{\ell^3} d\sigma' - \frac{5}{2r} \Delta g(\phi,\lambda,r), \quad (8.13)$$

with the definition

$$\kappa \stackrel{\text{def}}{=} 2 - \tfrac{3}{2} \frac{\left(r^2 - R^2\right)^2}{r^2 \ell^2}. \quad (8.14)$$

TABLEAU 8.2. Derivation of the kernel K' for the vertical gradient of gravity anomaly. Used is the definition of ℓ, equation 8.6, as well as Poisson's integral equation 8.10.

$$
\begin{aligned}
\frac{\partial \Delta g(\phi,\lambda,r)}{\partial r} &= \frac{1}{4\pi}\frac{\partial}{\partial r}\left(\iint_\sigma \left(\frac{R}{r}\cdot R\left(r^2-R^2\right)\cdot \ell^{-3}\right)\Delta g(\phi',\lambda',R)\,d\sigma'\right) = \\
&= \frac{R^2}{4\pi}\iint_\sigma \frac{\partial}{\partial r}\left(\frac{1}{r}\cdot\left(r^2-R^2\right)\cdot\ell^{-3}\right)\Delta g(\phi',\lambda',R)\,d\sigma' = \\
&= \frac{R^2}{4\pi}\iint_\sigma \left(\left(-\frac{r^2-R^2}{r^2}+\frac{2r}{r}\right)\cdot\frac{1}{\ell^3}+\frac{1}{r}\cdot\left(r^2-R^2\right)\cdot\frac{d\left(\ell^2\right)^{-3/2}}{d\left(\ell^2\right)}\cdot\frac{\partial\ell^2}{\partial r}\right)\Delta g(\phi',\lambda',R)\,d\sigma' = \\
&= \frac{R^2}{4\pi}\iint_\sigma \left(\frac{1}{\ell^3}\left(-\frac{r^2-R^2}{r^2}+2\right)+\frac{r^2-R^2}{r}\cdot\left(-\tfrac{3}{2}\ell^{-5}\right)\cdot\left(2r-2R\cos\psi\right)\right)\Delta g(\phi',\lambda',R)\,d\sigma' = \\
&= \frac{R^2}{4\pi}\iint_\sigma \frac{1}{\ell^3}\left(2-\tfrac{3}{2}\frac{r^2-R^2}{r}\frac{1}{\ell^2}\frac{\ell^2+r^2-R^2}{r}\right)\Delta g(\phi',\lambda',R)\,d\sigma' - \\
&\qquad -\frac{1}{r}\cdot\frac{1}{4\pi}\iint_\sigma \frac{R}{r}\frac{R\left(r^2-R^2\right)}{\ell^3}\Delta g(\phi',\lambda',R)\,d\sigma' = \\
&= \frac{R^2}{4\pi}\iint_\sigma \frac{1}{\ell^3}\left(2-\tfrac{3}{2}\frac{r^2-R^2}{r^2}-\tfrac{3}{2}\frac{\left(r^2-R^2\right)\left(r^2-R^2\right)}{r^2\ell^2}\right)\Delta g(\phi',\lambda',R)\,d\sigma' - \frac{1}{r}\Delta g(\phi,\lambda,r) = \\
&= \frac{R^2}{4\pi}\iint_\sigma \frac{1}{\ell^3}\left(2-\tfrac{3}{2}\frac{\left(r^2-R^2\right)^2}{r^2\ell^2}\right)\Delta g(\phi',\lambda',R)\,d\sigma' - \left(\frac{1}{r}+\frac{3}{2r}\right)\Delta g(\phi,\lambda,r) = \\
&= \frac{R^2}{4\pi}\iint_\sigma \frac{1}{\ell^3}\left[2-\tfrac{3}{2}\frac{\left(r^2-R^2\right)^2}{r^2\ell^2}\right]\Delta g(\phi',\lambda',R)\,d\sigma' - \frac{5}{2r}\Delta g(\phi,\lambda,r). \qquad (8.12)
\end{aligned}
$$

Looking at figure 8.7, it is seen that the Poisson kernel K gets narrower in proportion to $r-R$ and taller in proportion to $(r-R)^{-2}$. As the integral over the Poisson kernel is two-dimensional and scales with the square of the width, it remains constant when $r \to R$, and in fact the kernel converges to the two-dimensional Dirac δ function.

For the kernel K' of the vertical gradient of the gravity anomaly, the behaviour is more unpleasant: it narrows in the same way, but, as figure

8.7 shows, gets taller in proportion to $(r-R)^{-3}$. This makes its integral over the sphere diverge in proportion to $(r-R)^{-1}$.

Regularisation can be done by observing that a globally constant gravity anomaly field

$$\widetilde{\Delta g_0}(\phi,\lambda,r) = \widetilde{\Delta g_0}(r) = \left(\frac{R}{r}\right)^2 \Delta g_0$$

has a gradient of

$$\frac{\partial \widetilde{\Delta g_0}(\phi,\lambda,r)}{\partial r} = -\frac{2}{r}\widetilde{\Delta g_0}(\phi,\lambda,r), \tag{8.15}$$

but also, like equation 8.13:

$$\frac{\partial \widetilde{\Delta g_0}(\phi,\lambda,r)}{\partial r} = \frac{R^2}{4\pi}\iint_\sigma \kappa \frac{\widetilde{\Delta g_0}(\phi',\lambda',R)}{\ell^3}\,d\sigma' - \frac{5}{2r}\widetilde{\Delta g_0}(\phi,\lambda,r). \tag{8.16}$$

Subtract equation 8.16 from equation 8.13 and substitute equation 8.15, yielding

$$\begin{aligned}
\frac{\partial \Delta g(\phi,\lambda,r)}{\partial r} &= \frac{\partial\left(\Delta g(\phi,\lambda,r) - \widetilde{\Delta g_0}(\phi,\lambda,r)\right)}{\partial r} + \frac{\partial \widetilde{\Delta g_0}(\phi,\lambda,r)}{\partial r} = \\
&= \frac{R^2}{4\pi}\iint_\sigma \kappa \frac{\Delta g(\phi',\lambda',R) - \widetilde{\Delta g_0}(\phi',\lambda',R)}{\ell^3} d\sigma' - \\
&\qquad - \frac{5}{2r}\left(\Delta g(\phi,\lambda,r) - \widetilde{\Delta g_0}(\phi,\lambda,r)\right) - \frac{2}{r}\widetilde{\Delta g_0}(\phi,\lambda,r) = \\
&= \frac{R^2}{4\pi}\iint_\sigma \kappa \frac{\Delta g(\phi',\lambda',R) - \Delta g_0}{\ell^3} d\sigma' - \\
&\qquad - \frac{5}{2r}\left(\Delta g(\phi,\lambda,r) - \left(\frac{R}{r}\right)^2 \Delta g_0\right) - \frac{2}{r}\left(\frac{R}{r}\right)^2 \Delta g_0.
\end{aligned}$$

Choose the constant $\Delta g_0 \stackrel{\text{def}}{=} \Delta g(\phi,\lambda,R)$, the anomaly at sea level of the evaluation point:

$$\begin{aligned}
\frac{\partial \Delta g(\phi,\lambda,r)}{\partial r} &= \frac{R^2}{4\pi}\iint_\sigma \kappa \frac{\Delta g(\phi',\lambda',R) - \Delta g(\phi,\lambda,R)}{\ell^3} d\sigma' - \\
&- \frac{5}{2r}\left(\Delta g(\phi,\lambda,r) - \left(\frac{R}{r}\right)^2 \Delta g(\phi,\lambda,R)\right) - \frac{2}{r}\left(\frac{R}{r}\right)^2 \Delta g(\phi,\lambda,R) \approx
\end{aligned}$$

$$\approx \frac{R^2}{4\pi}\iint_\sigma \kappa \frac{\Delta g(\phi',\lambda',R)-\Delta g(\phi,\lambda,R)}{\ell^3}d\sigma' - \frac{2}{r}\left(\frac{R}{r}\right)^2 \Delta g(\phi,\lambda,R). \tag{8.17}$$

For $\kappa = 2$, this would correspond to Heiskanen and Moritz (1967) equation 2-217, however for an evaluation point at level $r \neq R$. For a well-behaved gravity-anomaly field,

$$\Delta g(\phi',\lambda',R) - \Delta g(\phi,\lambda,R) \overset{\text{nicely}}{\to} 0 \qquad \text{for } (\phi',\lambda') \to (\phi,\lambda),$$

and the integral 8.17 will converge for $r \to R$. We posit without proof that for $r \to R$, convergence will be to the same limit as the Heiskanen and Moritz equation, in other words, the second term in expression 8.14 fades away and, effectively, $\kappa \to 2$.

If we are integrating over the surface of a spherical Earth of radius R rather than the unit sphere σ of radius 1 — or, equivalently, in local metric co-ordinates (x, y) — we can make the substitution $dS = R^2\, d\sigma$, with dS a surface element on a sphere of radius R. This removes the factor R^2 from integral equations such as 8.10, 8.12, and 8.17.

In Molodensky's method this or similar equations can be rapidly evaluated from very local gravimetric data.

The closed expression given in Heiskanen and Moritz (1967, expression 2-217), is the anomalous vertical gravity gradient evaluated at sea level (on the reference sphere). In our equations 8.17 and 8.11 we also need gravity anomalies at sea level. However, anomalies at the *topographic surface* level are available. In practice, we may proceed iteratively, by initially assuming that the anomaly values observed at topography level are at sea level:

$$\Delta g^{(0)}(\phi,\lambda,R) \approx \Delta g(\phi,\lambda,r) = \Delta g(\phi,\lambda,R+H),$$

in which $H = H(\phi,\lambda)$ is the topographic height at point (ϕ,λ). When a crude anomalous gradient has been calculated, for example using equation 8.17, we may perform a real *reduction to sea level*, in linear

approximation:

$$\Delta g^{(1)}(\phi,\lambda,R) \approx \Delta g(\phi,\lambda,r) - \left.\frac{\partial \Delta g^{(0)}(\phi,\lambda,z)}{\partial z}\right|_{z=r} H.$$

This may be iterated.

8.7 Gravity reductions in geoid determination

8.7.1 *Classical methods*

Use of the Stokes equation for gravimetric geoid determination presupposes that all masses are *inside the geoid* — and that the exterior field is thus harmonic. For this reason we move the topographic masses computationally to inside the geoid, in a way that needs to be specified. The classical methods for this are

epäsuora vaikutus

- Helmert's (second) condensation method, section 6.5: the masses are shifted vertically down to the geoid into a surface density layer. After this, shifting measured gravity down from the topographic surface to sea level is easy. The indirect effect (the effect of the mass shifts on the geoid, the "restore" step) is small.
- Isostatic reduction, in which the effects of both the topography and its compensation, the "roots" of mountains below sea level, are computationally removed. The indirect effect of this method is larger. See section 6.7 and equation 6.8.
- Bouguer reduction, section 6.2: the effect of the topographic masses is brutally removed from the observed gravity data, and, after geoid calculation, it is equally brutally restored to the result. Bouguer anomalies contain large negative biases in the mountains and therefore, the indirect effect of Bouguer reduction is excessive and extends over a large area. This is why Bouguer reduction is used more rarely.

8.7.2 Downwards continuation in linear approximation

The approach described above can, following Molodensky, be *linearised*:

$$T = \underbrace{\frac{R}{4\pi}\iint_\sigma \left(\overbrace{\Delta g - \left.\frac{\partial \Delta g}{\partial z}\right|_{z=H'} H'}^{\Delta g^*(\phi',\lambda')} \right) S(\psi)\, d\sigma' + \left.\frac{\partial T}{\partial z}\right|_{z=H} H}_{T^*(\phi,\lambda)}. \qquad (8.18)$$

So, first we reduce the Δg measured and calculated at the topographic surface to sea level using the vertical gradient of the anomalies and the terrain height H' of the measurement point, with the result

$$\Delta g^*(\phi',\lambda') = \Delta g(\phi',\lambda',H') - \left.\frac{\partial \Delta g(\phi',\lambda',z)}{\partial z}\right|_{z=H'} H'.$$

After this, we apply, at sea level, the Stokes equation, and obtain the disturbing potential at sea level T^*. After this, the disturbing potential is "unreduced" back to terrain level, to the evaluation point, with the equation

$$T(\phi,\lambda,H) = T^*(\phi,\lambda) + \left.\frac{\partial T(\phi,\lambda,z)}{\partial z}\right|_{z=H} H.$$

In these equations T, its vertical derivative $\frac{\partial}{\partial H}T$, Δg, and *its* vertical derivative $\frac{\partial}{\partial H}\Delta g$ always belong to the *exterior* harmonic gravity field. The connection between them is the fundamental equation of physical geodesy, equation 5.5, in spherical geometry

fysikaalisen geodesian perusyhtälö

$$\Delta g = -\frac{\partial T}{\partial r} - \frac{2}{r}T, \qquad (5.6)$$

in which $r = R + H$. Here, we need firstly the vertical derivative of the disturbing potential. This is easy:

$$\frac{\partial T}{\partial H} = \frac{\partial T}{\partial r} = -\Delta g - \frac{2}{r}T,$$

where the first term on the right is directly measured, and the second term's T is obtained iteratively from the main product of the solution process.

Calculating the vertical gradient of gravity anomalies, that is the *anomalous vertical gradient of gravity*, is harder. For this task, section 8.6 offers calculation options. Luckily for practical calculations, the kernels of the integral equations are very localised and one does not need gravity anomalies from a very large area.

8.7.3 The evaluation point as the reference level

In the above equation 8.18 we used as the *reference level* the sea surface. This is arbitrary: we may use whatever reference level, for example H_0, in which case

$$T = \frac{R + H_0}{4\pi} \iint_\sigma \left(\Delta g - \left. \frac{\partial \Delta g}{\partial z} \right|_{z=H'} (H' - H_0) \right) S(\psi)\, d\sigma' + \\ + \left. \frac{\partial T}{\partial z} \right|_{z=H} (H - H_0)\,.$$

If we now choose $H_0 = H$, the last term drops off, and we obtain

$$T = \frac{R + H}{4\pi} \iint_\sigma \left(\Delta g - \left. \frac{\partial \Delta g}{\partial z} \right|_{z=H'} (H' - H) \right) S(\psi)\, d\sigma'.$$

In this case, the reduction takes place from the height of the Δg measurement point to the height of the T evaluation point. This is likely to be a shorter distance than from sea level to evaluation height, especially in the immediate surroundings of the evaluation point. This means that the *linearisation error will remain smaller*.[1] What is bad, on the other hand, is that the expression in parentheses is now different for each evaluation point. This complicates the use of FFT-based computation techniques, on which more later.

[1] The linearisation error could be even further tuned down by choosing as the evaluation level for the vertical gradient $z = \frac{1}{2}(H' + H)$.

Here, we were all the time discussing the determination of the *disturbing potential* $T(\phi, \lambda, H)$; this is in practice the same as determining the *height anomaly*

$$\zeta(\phi, \lambda, H) = \frac{T(\phi, \lambda, H)}{\overline{\gamma_{Hh}}} \approx \frac{T(\phi, \lambda, H)}{\gamma\left(\phi, \frac{1}{2}(H + h)\right)},$$

equation 7.1. Here, γ is normal gravity calculated for point latitude[2] ϕ and topographic height $\frac{1}{2}(H + h) \approx H + \frac{1}{2}\zeta$.

[2]

8.7.4 *The residual terrain modelling method*

Imagine that, conceptually, the topographic masses are shifted to below the geoid in a way that *does not change the exterior field.* This is materially the same as determining the geoid associated with the harmonically downwards continued exterior field.

The problem here is that such a mass distribution below sea level which produces the harmonically downwards continued external potential in the space between topographic surface and geoid does not always precisely exist. Or that a suitable mass distribution may exist but contains very large positive and negative masses close to each other, which is physically unrealistic.

One expresses this by saying that the problem is "*ill-posed*". In such cases, one uses *regularisation*: one changes the exterior field a little — as little as possible, so that it becomes a sensible field that *can* be harmonically continued below the topographic surface. Then, some sensible mass distribution interior to the geoid producing this field will also exist.

huonosti asetettu

One can start, for example, by filtering out the short-wave parts caused by the topography using a high-resolution digital terrain model. This is called the RTM (residual terrain modelling) method.

jäännösmaasto-mallinnus

[2] In an actual calculation one would calculate $\overline{\gamma_{Hh}}$ using the true geodetic latitude φ and equation 4.10. The height $\frac{1}{2}(H + h)$ has to be correct within a few metres in order to attain millimetre precision in ζ.

puskutraktori

In this method, we do not actually move all topographic masses to below the geoid. Instead, we use a bulldozer technique, figure 8.8: only masses close to the topographic surface are either removed or filled in, in a way that creates a smooth replacement topography that is long-wavelength only. The exterior field of this smoothed topography, unlike that of the original topography, lacks the shortest wavelengths. It may thus be downwards continued to the geoid with sufficient precision.

ilma-anomalia

First, we computationally remove from the topography *only* the short wavelengths (under 30 km) by moving the masses of the peaks into the valleys: a low-pass filtering. The effect of this on the free-air gravity anomalies Δg calculated from measurements is evaluated and taken into account: the "remove" step.

In detail:

1. At each point P we apply the *terrain correction* to the gravity anomalies as described in section 6.3.
2. Next, we remove the attraction of a Bouguer plate of thickness $H - H_{RTM}$, in which H stands for the terrain height of point P, and H_{RTM} for the height of the smoothed, or low-pass filtered, terrain at the horizontal location of point P. This effect is, according to equation 6.1, equal to

$$2\pi G\rho\,(H - H_{RTM})\,,$$

   in which ρ is the rock density assumed in the calculation.
3. After this, the *location* of the gravity anomaly is moved down (or up!) — "downwards continuation" — from the original terrain level H to the surface of the new, smoothed terrain, H_{RTM}. Equation 8.17 for the vertical gradient of the free-air gravity anomaly may be used for this.

   If this anomalous vertical gradient is ignored as it often is, then the vertical gravity gradient of the terrain-reduced external field is assumed to be the vertical gradient of normal gravity — according

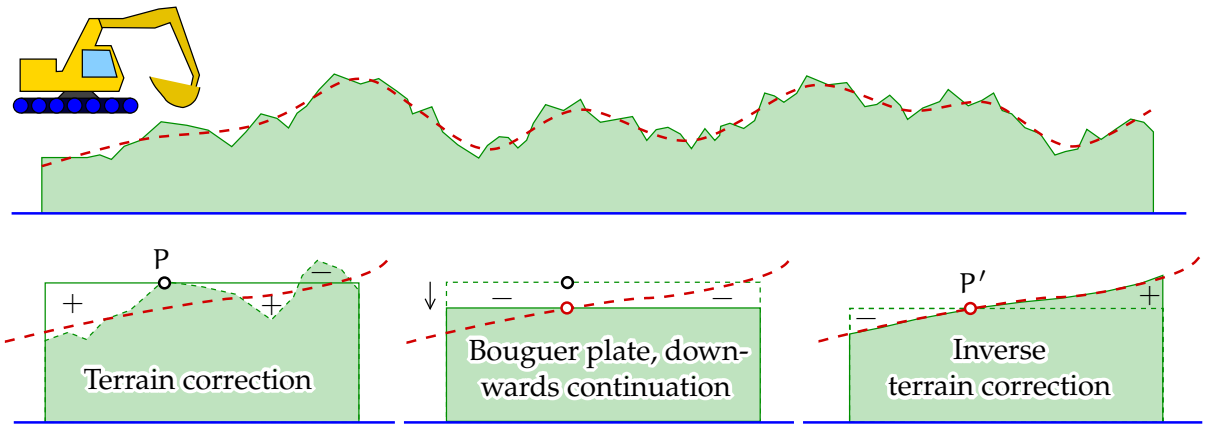


FIGURE 8.8. Residual terrain modelling (RTM). One removes the short wavelengths, the deviations from the red dashed line, from the terrain computationally: the masses rising above it are removed, the valleys below it are filled. After reduction, the red dashed line, smoother than the original terrain, is the new terrain surface. The exterior potential of the new mass distribution will differ only little from the original one, but may be harmonically downwards continued to sea level.
Left, terrain correction for point P, middle, Bouguer-plate and gradient reduction to the level of smoothed terrain point P′, and right, the inverse terrain correction for point P′.

to section 5.4, $-0.3\,^{\mathrm{mGal}}/_{\mathrm{m}}$ — and this operation will leave the gravity anomaly unchanged.

4. Rigorously speaking, an inverse terrain correction for the shapes of the smoothed terrain should be applied, to arrive at gravity anomalies realistic for this new replacement topography. Often also this step is left out as the effect is small.
5. After that, harmonic downwards continuation of the exterior field succeeds: almost only long wavelengths are left in the exterior field.

Because the mass shifts in the RTM method are so small, take place

over such small distances, and are of such a short wavelength in nature, the *indirect effect* or "restore" step — the change in geopotential due to the mass shifts that has to be applied in reverse to arrive at the final geopotential or geoid solution — is so small as to often be negligible. For the same reason, the effect of unknown topographic density will also remain small.

Finally we note that, because the RTM method removes the effect of the short-wavelength topography, it is also a suitable method for *interpolating* gravity anomalies. See Märdla (2017).

8.8 The remove–restore method

All current geoid determination methods are in one way or another "remove–restore" methods, even in several different ways.

poistamis-entistämis-menetelmä

pallofunktio-kehitelmä

1. From the observed gravity values, first the effect of a global gravity field model is removed. This model is generally given in the form of a spherical-harmonic expansion. Thus, a *residual gravity field* is obtained
   - that has numerically smaller values which are easier to work with
   - that is *more local*: the long "wavelengths", the patterns extending over large areas, have been removed from the residual field, only the local details remain.
2. From the observed gravity, the effects are removed of all masses that are *outside the geoid* — in practice, the topography. The purpose of this is to obtain a residual gravity field
   - to which the Stokes equation may be applied, because no masses are left outside the boundary surface
   - from which especially the very small "wavelengths" — details of the order of a few kilometres in size — caused by the topography, are gone. After this, *prediction* of gravity values

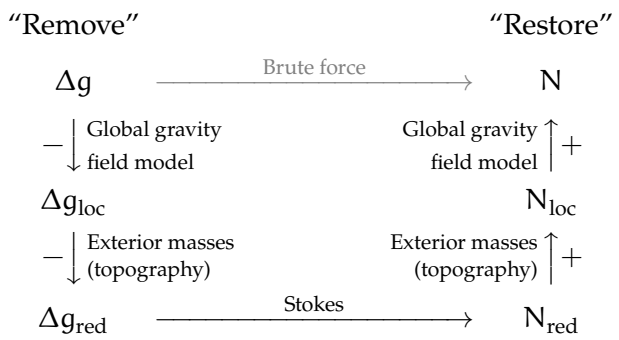


FIGURE 8.9. The remove–restore method as a commutative diagram.

from sparse measurement values will work better.

Some gravity reduction methods — methods which computationally remove the gravity effect of the exterior masses — with good prediction properties were already presented in subsection 8.7.1: Bouguer reduction and isostatic reduction. Also Helmert condensation may be mentioned, although its prediction properties are poorer.

We may illustrate the remove–restore method by means of *commutative diagram* 8.9. In this diagram, the black arrows with text denote calculations that are recommended, because they are easy and accurate. The grey arrow with text refers to direct computation, which again is computing intensive and numerically troublesome. kommutoiva kaavio

8.9 Kernel modification

In the remove–restore method described above, the handling of reduced gravity anomalies Δg_{red} and geoid heights N_{red} happens ordinarily within a small area. For example, when using the FFT method, the area of computation is often a rectangular area in the map projection plane, drawn generously around the country or area for which a geoid model is being computed.

Furthermore, if we compute a geoid model directly by integrating the

Stokes equation, we will evaluate this integral, after removing the effect of the global model from the given gravity data, only over a limited kalotti area or *cap*: evaluate the equation

$$N_{\text{red}} = \frac{R}{4\pi\gamma} \iint_{\sigma_0} S(\psi)\, \Delta g_{\text{red}}(\phi', \lambda')\, d\sigma', \tag{8.19}$$

in which σ_0 is a cap on the unit sphere the radius of which is, say, ψ_0.

The (possibly dangerous) assumption behind this is that, outside the cap, Δg_{red} is both small and rapidly varying, because the longer wavelengths have been removed from it with the global-model reduction.

Write both parts of the integrand in equation 8.19 into spectral form:

$$S(\psi) = \sum_{n=2}^{\infty} \frac{2n+1}{n-1} P_n(\cos\psi)$$

and

$$\Delta g_{\text{red}}(\phi', \lambda') = \sum_{n=L+1}^{\infty} \Delta g_n(\phi', \lambda'),$$

assuming that L is the largest degree number that is still along in the global spherical-harmonic expansion, or gravity model, that was subtracted from the data — and that the model is accurate up to that degree number.

Now, because Δg_n is a certain linear combination of the surface spherical harmonics

$$Y_{nm}(\psi, \alpha) = \begin{cases} P_{nm}(\cos\psi)\cos m\alpha & \text{if } m = 0, \ldots, n, \\ P_{n|m|}(\cos\psi)\sin|m|\,\alpha & \text{if } m = -n, \ldots, -1, \end{cases}$$

for example like this, compare equation 3.13:

$$\Delta g_n(\psi, \alpha) = \frac{1}{R^{n+1}} \sum_{m=-n}^{n} \Delta g_{nm} Y_{nm}(\psi, \alpha),$$

and also

$$P_n(\cos\psi) = P_{n0}(\cos\psi)\cos(0 \cdot \alpha) = Y_{n0}(\psi, \alpha),$$

it follows from the orthogonality of the Y functions that

$$\iint_\sigma P_n(\cos\psi)\,\Delta g_{n'}(\phi',\lambda')\,d\sigma' = 0 \qquad \text{if } n \neq n'.$$

Now we may write — the terms $n \leqslant L$ drop away:

$$\begin{aligned}
&\iint_\sigma S(\psi)\,\Delta g_{\text{red}}(\phi',\lambda')\,d\sigma' = \\
&\quad = \iint_\sigma \left(\sum_{n=2}^{\infty} \frac{2n+1}{n-1} P_n(\cos\psi)\right)\left(\sum_{n=L+1}^{\infty} \Delta g_n(\phi',\lambda')\right) d\sigma' = \\
&\quad = \iint_\sigma \left(\sum_{n=L+1}^{\infty} \frac{2n+1}{n-1} P_n(\cos\psi)\right)\left(\sum_{n=L+1}^{\infty} \Delta g_n(\phi',\lambda')\right) d\sigma' = \\
&\qquad\qquad = \iint_\sigma S^L(\psi)\,\Delta g_{\text{red}}(\phi',\lambda')\,d\sigma',
\end{aligned}$$

in which

$$S^L(\psi) = \sum_{n=L+1}^{\infty} \frac{2n+1}{n-1} P_n(\cos\psi)$$

is a so-called *modified Stokes kernel function*. The harmonic degree number L is called the *modification degree*. The size of the evaluation area σ_0 is chosen to be compatible with this.

The modification method described here, restricting the Legendre expansion of the S function to higher degree numbers, is called the *Wong–Gore*[3] modification (Wong and Gore, 1969). A desirable property of the new kernel function S^L is that it would be — at least compared to the original function S — *small* outside the cap area σ_0. In that case, restricting the integral to the cap instead of the whole unit sphere (equation 8.19) does not do much damage. It is clear that S^L is much narrower than S, as only the higher harmonic degrees are represented in it. This can be verified by plotting a graph of both curves (figure 8.10). The graph does not however go totally to zero outside the cap but oscillates somewhat.

[3] L. Wong and R. C. Gore worked at the Aerospace Corporation, a space technology research institution in California. Wikipedia, The Aerospace Corporation.

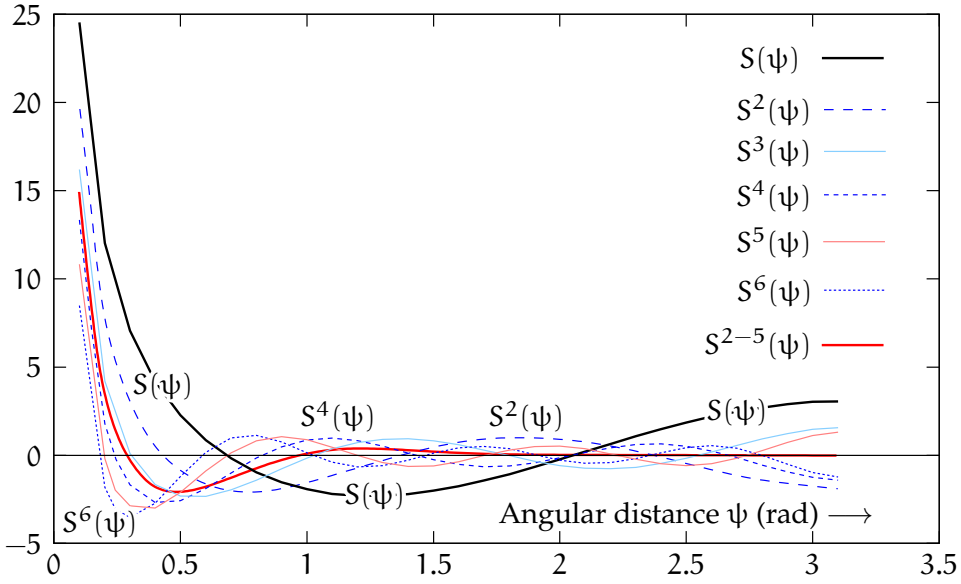


FIGURE 8.10. Modified Stokes kernel functions. Note how the kernel values for higher modification degree numbers L approach zero outside the local area. The red curve has been "soft modified" over a modification degree range of 2–5 using a cosine taper.

The reason for the oscillation is that in the frequency or degree-number domain the modified kernel's cut-off is quite sharp. Transforming such a sharp edge between the space and frequency domains will invariably produce an oscillation, which is related to the so-called *Gibbs*[4] *phenomenon*.

4 Gibbsin ilmiö

In figure 8.10 we have drawn in red a Stokes kernel that was modified or tapered "softly", by instead of removing the lower-degree terms altogether, forcing them gradually to zero going from degree number 5 down to 2. The curve is seen to go even better to zero than the "sharply" modified kernels.

[4]Josiah Willard Gibbs (1839–1903) was an American physicist, chemist, thermodynamicist, mathematician and engineer.

8.10 Advanced kernel modifications

Other kernel modification methods are found in the literature. Their general form is

$$S^L(\psi) = \sum_{n=L+1}^{\infty} \frac{2n+1}{n-1} P_n(\cos\psi) + \sum_{n=2}^{L} (1-s_n) \frac{2n+1}{n-1} P_n(\cos\psi) = \\ = S(\psi) - \sum_{n=2}^{L} s_n \frac{2n+1}{n-1} P_n(\cos\psi), \quad (8.20)$$

in which the modification coefficients s_n, $n = 2, \ldots, L$ can be chosen.[5] They are chosen so as to minimise the values of the kernel S^L in the area outside the cap, $\sigma - \sigma_0$. In this way one may eliminate the truncation error of equation 8.19 and the oscillation of the Wong–Gore modification almost entirely. Molodensky et al. (1962) had already developed such a method earlier. See also Bucha et al. (2019).

In the above equation 8.20 we want to minimise the function

$$S^L(\psi) = S(\psi) - \sum_{n=2}^{L} s_n \frac{2n+1}{n-1} P_n(\cos\psi)$$

over the area outside a local cap, $\sigma - \sigma_0$. Let us multiply this expression with each of the Legendre polynomials $P_n(\cos\psi)$, $n = 2, \ldots, L$ in turn, integrate over the area $\sigma - \sigma_0$ outside the local cap, and require the result to vanish:\{

$$\int_{\sigma-\sigma_0} S(\psi)\, P_n(\cos\psi)\, d\sigma - \\ - \sum_{n'=2}^{L} s_{n'} \frac{2n'+1}{n'-1} \int_{\sigma-\sigma_0} P_{n'}(\cos\psi)\, P_n(\cos\psi)\, d\sigma = 0, \\ n = 2, \ldots, L,$$

[5] The choice $s_n = 1$ again gives the simply (Wong–Gore) modified Stokes kernel from which the low degrees have been completely removed.

a system of $L-1$ equations in the $L-1$ unknowns $s_{n'}$:

$$\sum_{n'=2}^{L} \frac{2n'+1}{n'-1} e_{nn'} s_{n'} = Q_n,$$

in which

$$Q_n = \frac{1}{2\pi}\int_{\sigma-\sigma_0} S(\psi)\, P_n(\cos\psi)\, d\sigma = \int_{\psi_0}^{\pi} S(\psi)\, P_n(\cos\psi)\sin\psi\, d\psi$$

and

$$e_{nn'} = \frac{1}{2\pi}\int_{\sigma-\sigma_0} P_n(\cos\psi)\, P_{n'}(\cos\psi)\, d\sigma = \\ = \int_{\psi_0}^{\pi} P_n(\cos\psi)\, P_{n'}(\cos\psi)\sin\psi\, d\psi.$$

The coefficients Q_n are known as Molodensky's truncation coefficients, $e_{nn'}$ as Paul's (1973) coefficients.

From this, we can solve the s_n for every degree number n from 2 to L.

This solution also sets to zero the expressions

$$\left\langle S^L \cdot P_n \right\rangle_{\sigma-\sigma_0} = \int_{\sigma-\sigma_0} S^L(\psi)\, P_n(\cos\psi)\, d\sigma, \tag{8.21}$$

for all values n from 2 to L.

Expressions 8.21 can be understood as *inner or scalar products*, between functions S^L and P_n. Similarly, the elements of $e_{nn'}$ contain the scalar products between functions P_n and $P_{n'}$. These scalar products do not vanish: when integrating over $\sigma - \sigma_0$, unlike over the whole sphere σ, the Legendre polynomials are *not* mutually orthogonal. Therefore, e is a full matrix, not a diagonal matrix like when integrating over the full unit sphere σ.

The Legendre polynomials *are*, however, *independent* of each other on the domain $\sigma - \sigma_0$, and together span an $L-1$ -dimensional linear vector space.

Now, outside the cap σ_0 of radius ψ_0, the Stokes kernel $S(\psi)$, by visual inspection, is "smooth". Depending of course on the values of

cap radius ψ_0 and modification degree L, it may be so smooth that it does not contain any significant contribution from degree numbers higher than L. If this applies for S, it will also apply for S^L. This means that S^L will be a linear combination of the Legendre polynomials: an element of the vector space spanned by the polynomials P_n, $n = 2, \ldots,$ L. But if this is so, and the scalar products 8.21 with each of the basis vectors vanish, then S^L *must be the zero function* on $\sigma - \sigma_0$. kantavektori

See also Featherstone (2003).

Appendix A section A.1 explains more about linear vector spaces and the scalar product of vectors.

8.11 Block integration

In numerical gravimetric geoid determination one uses *averages* of anomalies computed over standard-sized cells or *blocks*, generally $5' \times 5'$, $10' \times 10'$, $30' \times 30'$ etcetera. At European latitudes, often sizes like $3' \times 5'$, $5' \times 10'$, $6' \times 10'$ are used, which are approximately square.

The following equation applies when evaluating an integral using block averages:

$$N(\phi,\lambda) \approx \frac{R}{4\pi\gamma} \sum_i S_i(\phi,\lambda)\, \overline{\Delta g}_i, \tag{8.22}$$

in which $\overline{\Delta g}_i$ is the mean of block i:

$$\overline{\Delta g}_i \stackrel{\text{def}}{=} \frac{1}{\omega_i} \iint_{\sigma_i} \Delta g(\phi,\lambda)\, d\sigma = \frac{1}{\omega_i} \iint_{\sigma_i} \Delta g(\phi,\lambda) \cos\phi\, d\phi\, d\lambda,$$

and the block integral of the Stokes kernel similarly

$$S_i(\phi,\lambda) \stackrel{\text{def}}{=} \iint_{\sigma_i} S\big(\psi(\phi,\lambda;\phi',\lambda')\big) \cos\phi'\, d\phi'\, d\lambda',$$

in which σ_i is the area of block i and its size on the unit sphere is

$$\omega_i \stackrel{\text{def}}{=} \iint_{\sigma_i} d\sigma = \iint_{\sigma_i} \cos\phi\, d\phi\, d\lambda.$$

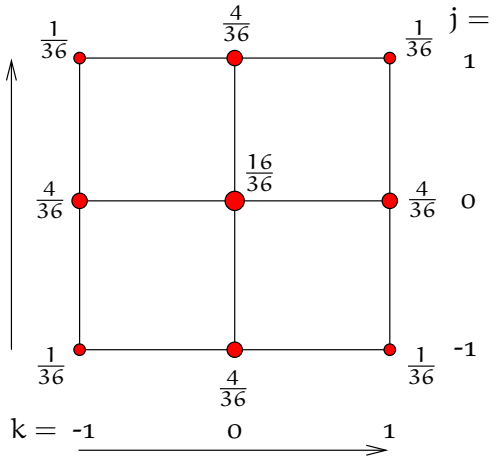


FIGURE 8.11. Simpson integration nodal weights in two dimensions.

Numerical evaluation of such an integral, or *quadrature*, is done conveniently using Simpson's rule:[6]

$$S_i(\phi,\lambda) = \int_{\lambda_i-\Delta\lambda/2}^{\lambda_i+\Delta\lambda/2} \int_{\phi_i-\Delta\phi/2}^{\phi_i+\Delta\phi/2} S\big(\psi(\phi,\lambda;\phi',\lambda')\big)\cos\phi'\,d\phi'\,d\lambda' \approx$$
$$\approx \Delta\phi\,\Delta\lambda \sum_{j=-1}^{1} w_j \sum_{k=-1}^{1} w_k S_i^{jk},$$

in which $\Delta\lambda$ and $\Delta\varphi$ are the block sizes in the latitude and longitude directions, and $w_{-1} = w_1 = \frac{1}{6}$, $w_0 = \frac{4}{6}$ are the weights.

$$S_i^{jk}(\phi,\lambda) = S\Big(\psi\big(\phi,\lambda;\phi_i + \tfrac{1}{2}j\,\Delta\phi, \lambda_i + \tfrac{1}{2}k\,\Delta\lambda\big)\Big)\cos\big(\phi_i + \tfrac{1}{2}j\,\Delta\phi\big),$$
$$j,k = -1,0,1$$

are the values of expression $S\big(\psi(\phi,\lambda;\phi',\lambda')\big)\cos\phi'$ at the nodal points used in the evaluation, 3×3 of them. See figure 8.11. More complicated formulas (repeated Simpson or Romberg) can also be employed.

[6] Thomas Simpson FRS (1710–1761) was an English mathematician and textbook writer. Actually Simpson's rule was already being used a century earlier by Johannes Kepler.

8.12 Effect of the local zone

One can show that the effect of the local (inner) zone on the geoid height at the evaluation point (ϕ, λ) is proportional to the gravity anomaly in the point itself, $\Delta g(\phi, \lambda)$. Starting from Stokes equation 8.2 with $S(\psi) \approx {}^{1}/_{\sin \frac{1}{2}\psi} \approx {}^{2}/_{\psi}$, we find, for a circular inner zone of radius ψ_0:

$$\delta N_0 = \frac{R}{4\pi\gamma} \int_0^{2\pi} \int_0^{\psi_0} \frac{2}{\psi} \Delta g(\psi, \alpha) \cancel{\sin\psi}\, d\psi\, d\alpha \approx$$
$$\approx \frac{R}{\gamma} \int_0^{\psi_0} \left(\frac{1}{2\pi} \int_0^{2\pi} \Delta g(\psi, \alpha)\, d\alpha \right) d\psi \approx \frac{R}{\gamma} \cdot \psi_0 \cdot \overline{\Delta g_0} = \frac{s_0}{\gamma} \overline{\Delta g_0}.$$

Here $s_0 = R\psi_0$ is the radius of the local block or cap in units of length. The quantity

$$\overline{\Delta g_0} \stackrel{\text{def}}{=} \frac{1}{\psi_0} \int_0^{\psi_0} \left(\frac{1}{2\pi} \int_0^{2\pi} \Delta g(s, \alpha)\, d\alpha \right) d\psi =$$
$$= \frac{1}{s_0} \int_0^{s_0} \left(\frac{1}{2\pi} \int_0^{2\pi} \Delta g(s, \alpha)\, d\alpha \right) ds$$

is a special average of the gravity anomaly, the average of "ring averages" for radii s between zero and s_0. If s_0 is small, one may take for this the anomaly value $\Delta g(\phi, \lambda)$ at the centre without incurring much error.

The local contributions to the deviations of the plumb line are again proportional to the *horizontal gradients* of gravity anomalies. We start from Vening Meinesz equations 8.5, with the above approximation for a local cap, specifically

$$S(\psi) \approx \frac{2}{\psi} \quad \Longrightarrow \quad \frac{d}{d\psi} S(\psi) = -\frac{2}{\psi^2} :$$
$$\begin{Bmatrix} \delta\xi_0 \\ \delta\eta_0 \end{Bmatrix} \approx \frac{1}{4\pi\gamma} \int_0^{\psi_0} \int_0^{2\pi} \left(-\frac{2}{\psi^2} \right) \Delta g(\phi', \lambda') \begin{Bmatrix} \cos\alpha \\ \sin\alpha \end{Bmatrix} \sin\psi\, d\alpha\, d\psi.$$

We expand Δg into local linear rectangular co-ordinates x, y:

$$\Delta g \approx \Delta g_0 + x \frac{\partial \Delta g}{\partial x} + y \frac{\partial \Delta g}{\partial y} \approx$$

$$\approx \Delta g_0 + R\psi\left(\cos\alpha\frac{\partial\Delta g}{\partial x} + \sin\alpha\frac{\partial\Delta g}{\partial y}\right),$$

and substitute:

$$\left\{\begin{array}{c}\delta\xi_0\\ \delta\eta_0\end{array}\right\} \approx \frac{1}{4\pi\gamma}\cdot$$
$$\cdot\int_0^{\psi_0}\int_0^{2\pi} -\frac{2}{\psi^2}\left(\Delta g_0 + R\psi\left(\cos\alpha\frac{\partial\Delta g}{\partial x} + \sin\alpha\frac{\partial\Delta g}{\partial y}\right)\right)\left\{\begin{array}{c}\cos\alpha\\ \sin\alpha\end{array}\right\}\sin\psi\,d\alpha\,d\psi.$$

Here, the terms in Δg_0 drop out in α integration, because $\int_0^{2\pi}\sin\alpha\,d\alpha = \int_0^{2\pi}\cos\alpha\,d\alpha = 0$. So do the mixed terms in $\sin\alpha\cos\alpha$. The only nonzero terms contain $\int_0^{2\pi}\sin^2\alpha\,d\alpha = \int_0^{2\pi}\cos^2\alpha\,d\alpha = \pi$:

$$\begin{aligned}\delta\xi_0 &\approx -\frac{1}{4\pi\gamma}\int_0^{\psi_0}\int_0^{2\pi}\frac{2}{\cancel{\psi^2}}R\cancel{\psi}\cos\alpha\frac{\partial\Delta g}{\partial x}\cos\alpha\,\cancel{\sin\psi}\,d\alpha\,d\psi \approx\\ &\approx -\frac{R}{2\pi\gamma}\int_0^{\psi_0}\int_0^{2\pi}\frac{\partial\Delta g}{\partial x}\cos^2\alpha\,d\alpha\,d\psi \approx -\frac{R\psi_0}{2\gamma}\frac{\partial\Delta g}{\partial x},\\ \delta\eta_0 &\approx -\frac{1}{4\pi\gamma}\int_0^{\psi_0}\int_0^{2\pi}\frac{2}{\cancel{\psi^2}}R\cancel{\psi}\sin\alpha\frac{\partial\Delta g}{\partial y}\sin\alpha\,\cancel{\sin\psi}\,d\alpha\,d\psi \approx\\ &\approx -\frac{R}{2\pi\gamma}\int_0^{\psi_0}\int_0^{2\pi}\frac{\partial\Delta g}{\partial y}\sin^2\alpha\,d\alpha\,d\psi \approx -\frac{R\psi_0}{2\gamma}\frac{\partial\Delta g}{\partial y}.\end{aligned}$$

Evaluating these integrals assumes the partial derivatives to be constant within the cap. Using $R\psi_0 = s_0$ yields now

$$\delta\xi_0 \approx -\frac{s_0}{2\gamma}\frac{\partial\Delta g}{\partial x},\quad \delta\eta_0 \approx -\frac{s_0}{2\gamma}\frac{\partial\Delta g}{\partial y}.$$

These equations might be useful as standard block integration, equation 8.22, behaves numerically poorly in the immediate surroundings of the evaluation point if the kernel function is singular at the origin $\psi = 0$. Both the Stokes 8.2 and Vening Meinesz 8.5 kernels are of this kind.

Self-test questions

1. What do the Stokes equation and its spectral form look like?

2. What does the Stokes kernel function $S(\psi)$ look like when expanded in Legendre polynomials?
3. What is a suitable approximation of the Stokes kernel when ψ is small?
4. What is an isotropic, what an anisotropic quantity on the Earth's surface? Give an example of the latter.
5. What does the Poisson integral equation describe?
6. Why are gravity reductions necessary when using the Stokes equation for computing a geoid model?
7. Which gravity reduction methods are available?
8. Explain the residual terrain modelling (RTM) method.
9. Explain the remove–restore approach.
10. Why, in geoid determination, is the Stokes kernel function often modified? What does such a modification look like?
11. What is the Gibbs phenomenon?

Exercise 8–1: The Stokes equation in the near zone

1. Derive a simpler form of the Stokes function $S(\psi)$ which is valid when the angular distance ψ is *small*. This simpler form really consists of only one term!
2. Using this form, write the integral equation

$$N = \frac{R}{4\pi\gamma} \iint_\sigma S(\psi)\, \Delta g \, d\sigma$$

into polar co-ordinates, as an integral of the form

$$\int_0^{2\pi} \int_0^{\infty} \cdots \, ds \, d\alpha,$$

in which $s = \psi R$ is the linear distance from the evaluation point, and α the azimuth angle (direction angle) from the evaluation point for the geoid height N to the moving data point for the gravity anomaly Δg.

Hint: you need to consider *Jacobi's determinant* for the polar coordinates (s, α).

3. Compute N (as an equation) if $\Delta g = \Delta g_0$ only within a circular area $s \leqslant s_0$, and outside it $\Delta g = 0$. Assume that s_0 is small.

Spectral techniques, FFT

9

9.1 The Stokes equation as a convolution

We start from the Stokes equation 8.1,

$$T(\phi,\lambda) = \frac{R}{4\pi}\iint_{\sigma} S(\psi)\,\Delta g(\phi',\lambda')\,d\sigma',$$

in which (ϕ',λ') is the location of the moving integration or observation point, and (ϕ,λ) is the location of the evaluation point, both at sea level, on the surface of a spherical Earth. So, the locations of both points are given in spherical co-ordinates. The integration is carried out over the surface of the unit sphere σ: a surface element is $d\sigma = \cos\phi\,d\phi\,d\lambda$, in which $\cos\phi$ is the determinant of Jacobi, for the spherical co-ordinates (ϕ,λ).

However *locally*, in a sufficiently small area, one may also write the point co-ordinates in rectangular form and express the integral in rectangular co-ordinates. Suitable rectangular co-ordinates are, for example, *map projection co-ordinates*, see figure 9.1.

A simple example of rectangular co-ordinates in the tangent plane would be

$$x = \psi R\cos\alpha, \quad y = \psi R\sin\alpha, \tag{9.1}$$

in which α is the azimuth of the line connecting evaluation point and moving data point. The centre of this projection is the point where the

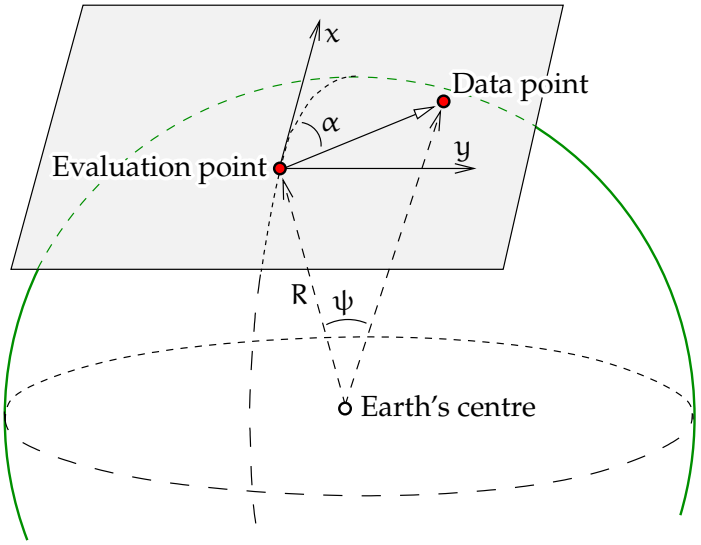


FIGURE 9.1. Map projection co-ordinates x, y in the local tangent plane.

tangent plane touches the sphere. The locations of other points are measured by the angle ψ at the Earth's centre, the *geocentric angular distance*, and by the direction angle in the tangent plane or *azimuth* α.

A more realistic example uses a popular conformal map projection called the *stereographic projection*:

$$x = 2\tan\frac{\psi}{2}\,R\cos\alpha, \quad y = 2\tan\frac{\psi}{2}\,R\sin\alpha.$$

In the limit for small values of ψ this agrees with equations 9.1.

Taking the squares of equations 9.1, summing them, and dividing the result by R^2 yields

$$\psi^2 \approx \frac{x^2+y^2}{R^2}.$$

More generally ψ is the angular distance between the points (x, y) (evaluation point) and (x', y') (data, integration or moving point) seen from the Earth's centre, approximately

$$\psi^2 \approx \left(\frac{x-x'}{R}\right)^2 + \left(\frac{y-y'}{R}\right)^2.$$

Furthermore, we must account for Jacobi's determinant R^2 of the projection:

$$d\sigma = R^{-2}\, dx\, dy \iff dx\, dy = R^2\, d\sigma,$$

and the Stokes equation now becomes

$$T(x,y) \approx \frac{1}{4\pi R} \iint_{-\infty}^{\infty} S(x-x', y-y')\, \Delta g(x', y')\, dx'\, dy', \qquad (9.2)$$

a two-dimensional *convolution*.[1]

Convolutions have nice properties in Fourier theory. If we designate the Fourier transform with the symbol $\mathcal{F}$, and convolution with the symbol $\otimes$, we may abbreviate the above equation as follows:

$$T = \frac{1}{4\pi R} S \otimes \Delta g,$$

and according to the *convolution theorem* ("Fourier transforms a convolution into a multiplication"):

$$\mathcal{F}\{T\} = \frac{1}{4\pi R} \mathcal{F}\{S\} \cdot \mathcal{F}\{\Delta g\}.$$

This approximation in the (x,y) plane works only *if integration can be restricted to a local area*, where the curvature of the Earth's surface may be neglected. This is possible thanks to the use of global spherical-harmonic expansions, because these represent the long-wavelength part of the spatial variability of the Earth's gravity field. After we have removed the effect of the global spherical-harmonic model from the observed gravity anomalies Δg (the "remove" step) we may safely forget the effect of areas far removed from the evaluation point: after this removal, the anomaly field Δg_{loc} will contain only the remaining short-wavelength parts, the effect of which cancels out over greater distances.

pallofunktiokehitelmä

poistamisvaihe

Of course, once the integral has been computed and the local disturb- häiriöpotentiaali

[1] The integration extends from minus to plus infinity in both co-ordinates x and y. This can only be realistic on a curved Earth if it is assumed that the kernel S is of *bounded support*: it differs from zero only in a bounded area, a small part of the whole Earth's surface. This is the case for the modified kernels discussed in section 8.9.

ing potential T_{loc}, and the corresponding geoid undulation N_{loc}, have been obtained, we must remember to add to them again the effect of the global spherical-harmonic expansion on the disturbing potential T and geoid undulation N to be calculated separately. This is the "restore" step of the computation; see the commutative diagram 8.9.

entistämisvaihe

kommutoiva kaavio

9.2 Integration by FFT

The Fourier transform needed for applying the convolution theorem is calculated as a *discrete Fourier transform*. The highly efficient Fast Fourier Transform, FFT, exists for this purpose, for example Vermeer (1993). There are multiple slightly differing formulations of the discrete Fourier transform to be found in the literature. It does not really matter which is chosen, as long as it is a compatible pair of a forward Fourier transform $\mathcal{F}$ and a reverse Fourier transform $\mathcal{F}^{-1}$.

hilaesitys

In preparation for this, we first build a discrete *grid representation* of the function $\Delta g(x, y)$, a rectangular table of Δg values on an equispaced (x_i, y_j) grid of points. The values may be, say, the function values themselves at the grid points:[2]

[2]

$$\Delta g_{ij} = \Delta g(x_i, y_j),$$

in which the co-ordinates of the grid points are

$$x_i = i\,\delta x, \quad y_j = j\,\delta y, \qquad i, j = 0, 1, \ldots, N-1,$$

for suitably chosen grid spacings $(\delta x, \delta y)$. The integer N is the grid size, assumed for simplicity to be the same in both directions.

Next, we do the same for the kernel function

$$S(\psi) = S(x - x', y - y') = S(\Delta x, \Delta y),$$

[2] Alternatively, one could for example calculate for every grid point the average over a square cell surrounding the point.

so we write

$$S_{ij} = S(\Delta x_i, \Delta y_j),$$

where again

$$\Delta x_i = i\,\delta x, \quad \Delta y_j = j\,\delta y, \qquad i, j = 0, 1, \ldots, N-1.$$

Now the central peak at the origin of the kernel function S — $S(\Delta x, \Delta y) \to \infty$ when $(\Delta x, \Delta y) \to (0, 0)$ — is placed at the origin $i = j = 0$ of the grid of function values S_{ij}, in one corner, and the grid contains only one quadrant of the peak. This is not acceptable.

The periodicity inherent in the discrete Fourier transform means that the values $i = \frac{1}{2}N, \ldots, N-1$ may be replaced by the negative values $i' \stackrel{\text{def}}{=} i - N = -\frac{1}{2}N, \ldots, -1$ without formally changing anything: see footnote 1 in appendix C. In this interpretation

$$\Delta x_{i'} = i'\,\delta x, \quad \Delta y_{j'} = j'\,\delta y, \qquad i', j' = -\tfrac{1}{2}N, \ldots, \tfrac{1}{2}N - 1,$$

landing the origin in the centre of the grid. This is the correct way to compute the values of the true, non-periodic kernel, with both positive and negative values Δx and Δy from an area symmetrically surrounding the origin.

Next:

1. The grid representations Δg_{ij} and S_{ij} thus obtained of the functions Δg and S are transformed to the *frequency domain* — they become functions $\mathcal{S}_{uv} = \mathcal{F}\{S_{ij}\}$ and $\mathcal{G}_{uv} = \mathcal{F}\{\Delta g_{ij}\}$ of the two "frequencies", the wave indices u and v in the x and y directions. The spatial frequencies or wave numbers[3] $\tilde{\nu}$ and spatial wavelengths λ are $\tilde{\nu}_x = \lambda_x^{-1} = u/L$, $\tilde{\nu}_y = \lambda_y^{-1} = v/L$, in which $L = N\,\delta x = N\,\delta y$ is the size of the area, assumed to be square.

[3] This is the so-called linear wave number, whole waves per unit of length. The angular wave number is $k = 2\pi\tilde{\nu}$, radians of phase angle per unit of length.

2. They are multiplied with each other "one frequency pair at a time": we calculate

$$\mathcal{T}_{uv} = \frac{1}{4\pi R}\, \mathcal{S}_{uv} \cdot \mathcal{G}_{uv}, \qquad u, v = 0, 1, \ldots, N-1. \tag{9.3}$$

3. We transform the result, $\mathcal{T}_{uv} = \mathcal{F}\{T_{ij}\}$, back to the space domain: $T_{ij} = \mathcal{F}^{-1}\{\mathcal{T}_{uv}\}$, a point grid $T_{ij} = T(x_i, y_j)$ of the disturbing potential T. The disturbing potential of an arbitrary point can be obtained from this grid by interpolation. The co-ordinates x_i, y_j run as functions of i, j in the same way as described above for Δg.

The method described is good for computing the disturbing potential T — and similarly the geoid height $N = T/\gamma$ — from gravity anomalies using the Stokes equation. It is just as good for evaluating other quantities, like for example the vertical gradient of the gravity anomaly using equation 8.17. The only requirement is that the equation can be expressed as a *convolution*.

Inversion calculation is also easy, as we shall see: in the frequency domain it is just a simple division.

Using the discrete Fourier transform requires that the input data, the field to be integrated — in the example, gravity anomalies — is given on a regular grid covering the area of computation, or can be converted into one. The result — in the example, the disturbing potential — is obtained on a regular grid in the same geometry. Values can then be *interpolated* to chosen locations.

The FFT method may be depicted as a *commutative diagram*, figure 9.2.

Appendix C offers a short explanation of why FFT works and what makes it as efficient as it is.

9.3 Solution in latitude and longitude

In the above equation 9.2, the grid co-ordinates x and y are rectangular. For practical reasons, we would rather use latitude and longitude (φ, λ) as grid co-ordinates. In that way, the need to generate a new (x, y) point

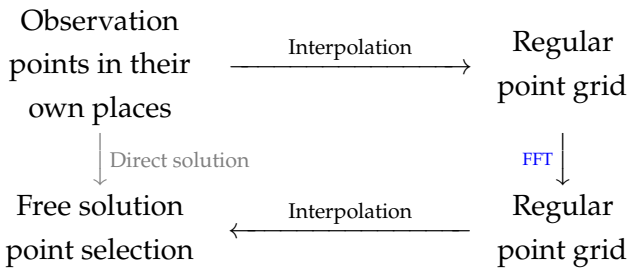


FIGURE 9.2. Commutative diagram for FFT.

grid by interpolating from the given (φ, λ) one through a map projection calculation is avoided. However, working in geographical co-ordinates causes errors due to *meridian convergence* — as a latitude and longitude co-ordinate system is not actually rectangular. The co-ordinate pair $(\varphi, \lambda \cos \varphi)$ would be slightly more suitable.

The problem has also been addressed on a more conceptual level.

9.3.1 *The Strang van Hees method*

The Stokes kernel function $S(\psi)$ depends only on the geocentric angular distance ψ between evaluation point (ϕ, λ) and observation point (ϕ', λ'). The angular distance may be written as follows (cosine rule on the sphere):

$$\cos \psi = \sin \phi \sin \phi' + \cos \phi \cos \phi' \cos(\lambda - \lambda').$$

Substitute

$$\begin{aligned} \cos(\lambda - \lambda') &= 1 - 2 \sin^2 \frac{\lambda - \lambda'}{2}, \\ \cos \psi &= 1 - 2 \sin^2 \frac{\psi}{2}, \\ \cos(\phi - \phi') &= 1 - 2 \sin^2 \frac{\phi - \phi'}{2}, \end{aligned}$$

and obtain the *half-angle cosine rule*:

$$\cos\psi = \cos(\phi - \phi') - 2\cos\phi\cos\phi'\sin^2\frac{\lambda-\lambda'}{2}$$
$$\implies \quad \sin^2\frac{\psi}{2} = \sin^2\frac{\phi-\phi'}{2} + \cos\phi\cos\phi'\sin^2\frac{\lambda-\lambda'}{2}.$$

Here we may use the following approximation:

$$\cos\phi', \cos\phi \approx \cos\phi_0,$$

vertaus-leveysaste

in which ϕ_0 is a reference latitude in the middle of the calculation area. Now the above equation becomes

$$\sin^2\frac{\psi}{2} \approx \sin^2\frac{\phi-\phi'}{2} + \cos^2\phi_0\sin^2\frac{\lambda-\lambda'}{2}, \tag{9.4}$$

which depends only on the differences $\Delta\phi \overset{\text{def}}{=} \phi - \phi'$ and $\Delta\lambda \overset{\text{def}}{=} \lambda - \lambda'$.

After this, the FFT method may be applied by using co-ordinates [4] (ϕ, λ)[4] and the Stokes kernel written as

$$S(\psi) = S(\Delta\phi, \Delta\lambda) = S\left(2\arcsin\sqrt{\sin^2\frac{\Delta\phi}{2} + \cos^2\phi_0\sin^2\frac{\Delta\lambda}{2}}\right),$$

which is now a function of only the differences $\Delta\phi$ and $\Delta\lambda$, as the convolution theorem requires. This clever way of using FFT in geographical [5] co-ordinates was invented by the Dutchman G. Strang van Hees[5] in 1990.

9.3.2 "Spherical FFT", multi-band model

We divide the area into several narrow latitude bands. In each band we apply the Strang van Hees method using its own optimal central latitude.

[4] In practice one uses the geodetic or geographical latitude φ instead of ϕ without significant error.

[5] Govert L. Strang van Hees (1932–2012) was a Dutch gravimetric geodesist.

Write the Stokes equation as follows:

$$N(\phi,\lambda) = \frac{R}{4\pi\gamma}\iint S(\Delta\phi,\Delta\lambda;\phi)\left[\Delta g(\phi',\lambda')\cos\phi'\right]d\phi'\,d\lambda', \qquad (9.5)$$

where we have expressed S as a function of latitude difference, longitude difference and *evaluation latitude*. Now, choose two support latitudes, ϕ_i and ϕ_{i+1}. Assume furthermore that between these S is a sufficiently linear function of ϕ. In that case we may write

$$S(\Delta\phi,\Delta\lambda;\phi) = \frac{(\phi-\phi_i)\,S_{i+1}(\Delta\phi,\Delta\lambda) + (\phi_{i+1}-\phi)\,S_i(\Delta\phi,\Delta\lambda)}{\phi_{i+1}-\phi_i},$$

where $\Delta\phi = \phi - \phi'$, $\Delta\lambda = \lambda - \lambda'$, and

$$S_i(\Delta\phi,\Delta\lambda) = S(\phi-\phi',\lambda-\lambda';\phi_i),$$
$$S_{i+1}(\Delta\phi,\Delta\lambda) = S(\phi-\phi',\lambda-\lambda';\phi_{i+1}).$$

We obtain by substitution into integral equation 9.5:

$$N(\phi,\lambda) = \frac{R}{4\pi\gamma}\left(\frac{\phi_{i+1}-\phi}{\phi_{i+1}-\phi_i}I_i + \frac{\phi-\phi_i}{\phi_{i+1}-\phi_i}I_{i+1}\right), \qquad (9.6)$$

with

$$I_i = \iint S_i(\Delta\phi,\Delta\lambda)\left[\Delta g(\phi',\lambda')\cos\phi'\right]d\phi'\,d\lambda',$$
$$I_{i+1} = \iint S_{i+1}(\Delta\phi,\Delta\lambda)\left[\Delta g(\phi',\lambda')\cos\phi'\right]d\phi'\,d\lambda'.$$

Equation 9.6 is the linear combination of two convolutions. Both are evaluated by FFT. The equation forms the weighted mean from the solutions obtained.

In this method we use, instead of approximative equation 9.4, an exact equation in which ϕ' is expressed into ϕ and $\Delta\phi$:

$$\sin^2\frac{\psi}{2} = \sin^2\frac{\phi-\phi'}{2} + \cos\phi\cos\phi'\sin^2\frac{\lambda-\lambda'}{2} =$$
$$= \sin^2\frac{\Delta\phi}{2} + \cos\phi\cos(\phi-\Delta\phi)\sin^2\frac{\Delta\lambda}{2}.$$

We calculate S_i and S_{i+1} for the support latitude values ϕ_i and ϕ_{i+1}, we evaluate the integrals with the aid of the convolution theorem, and

interpolate $N(\phi,\lambda)$ according to equation 9.6 when $\phi_i \leqslant \phi < \phi_{i+1}$. After this, the solution is not exact, because inside every band we still use linear interpolation. However by making the bands narrower, we can keep the error arbitrarily small.

9.3.3 *"Spherical FFT", Taylor expansion model*

This somewhat more complicated but also more versatile approach expands the Stokes kernel into a Taylor series with respect to latitude about a *reference latitude* located in the middle of the computation area.[6] Each term in the expansion depends only on the *difference* in latitude. The integral to be calculated similarly expands into terms, of which each contains a pure convolution.

Let us write the *general* problem as follows:

$$\ell(\phi,\lambda) = \int_0^{2\pi} \int_{-\pi/2}^{+\pi/2} C\left(\phi,\phi',\Delta\lambda\right)\left[m(\phi',\lambda')\cos\phi'\right] d\phi'\, d\lambda',$$

in which ℓ contains values to be computed, m values given, and C is the coefficient or kernel function. Here only *rotational symmetry* around the Earth's axis is assumed for the geometry: the kernel function depends only on the difference between longitudes $\Delta\lambda$ rather than the absolute longitudes λ and λ'.

In a concrete case, m contains for example gravity anomaly values Δg in various points (ϕ',λ'), ℓ contains geoid heights N in various points (ϕ,λ), and C contains coefficients calculated using the Stokes kernel function.

We first change the dependence upon ϕ and ϕ' into a dependence upon ϕ and $\Delta\phi$:

$$C = C\left(\phi,\phi',\Delta\lambda\right) = C(\Delta\phi,\Delta\lambda;\phi).$$

[6] In the literature the method has been generalised by also expanding the kernel with respect to height.

Linearise:

$$C = C_0(\Delta\phi, \Delta\lambda) + (\phi - \phi_0)\, C_\phi(\Delta\phi, \Delta\lambda) + \cdots$$

where we define for a suitable *reference latitude* ϕ_0:

$$\begin{aligned} C_0(\Delta\phi, \Delta\lambda) &\stackrel{\text{def}}{=} C\big(\Delta\phi, \Delta\lambda; \phi_0\big), \\ C_\phi(\Delta\phi, \Delta\lambda) &\stackrel{\text{def}}{=} \frac{\partial}{\partial\phi} C(\Delta\phi, \Delta\lambda; \phi)\Big|_{\phi=\phi_0}. \end{aligned}$$

This expansion into two terms will be accurate only for a limited range in $\Delta\varphi$, and the kernel function C is assumed to be of bounded support. In this case, the integrals may be calculated within a limited area instead of over the whole Earth.

Substitution yields

$$\begin{aligned} \ell(\phi, \lambda) &= \iint C(\Delta\phi, \Delta\lambda; \phi) \cdot m\big(\phi', \lambda'\big) \cos\phi' \, d\phi' \, d\phi' = \\ &= \iint \big(C_0 + (\phi - \phi_0)\, C_\phi\big) \cdot m \cos\phi' \, d\phi' \, d\phi' = \\ &= \iint C_0 \cdot m \cos\phi' \, d\phi' \, d\lambda' + (\phi - \phi_0) \iint C_\phi \cdot m \cos\phi' \, d\phi' \, d\lambda'. \end{aligned} \tag{9.7}$$

It is important here now that the integrals in the first and second terms,

$$\begin{aligned} \iint C_0(\Delta\phi, \Delta\lambda) \Big[m\big(\phi', \lambda'\big) \cos\phi' \Big] \, d\phi' \, d\lambda' &= C_0 \otimes \big[m \cos\phi \big], \\ \iint C_\phi(\Delta\phi, \Delta\lambda) \Big[m\big(\phi', \lambda'\big) \cos\phi' \Big] \, d\phi' \, d\lambda' &= C_\phi \otimes \big[m \cos\phi \big], \end{aligned}$$

are both convolutions: both C functions depend only on $\Delta\phi$ and $\Delta\lambda$. Both integrals can be calculated if both the data grid $m \cos\phi$ and the coefficient grids C_0 and C_ϕ are calculated first in preparation. After this — in principle expensive, but, thanks to FFT and the convolution theorem, a lot cheaper — integration, computing compound 9.7 is fast: one multiplication and one addition for each evaluation point (ϕ, λ).

Example Let the evaluation area at latitude 60° be 10° × 20° in size. If the grid mesh size is 5′ × 10′, the number of cells is 120 × 120.

Let us choose, say, a 256×256 size grid (so $N = 256$) and fill in the missing values by extrapolation.

The values of the kernel functions C_0 and C_ϕ are calculated on a 256×256 grid $(\Delta\phi, \Delta\lambda)$ as well. The number of these is thus also 65 536. Calculating the convolutions $C_0 \otimes [m \cos\phi]$ and
7 $C_\phi \otimes [m \cos\phi]$ by means of FFT — like[7]

$$\iint C_0(\Delta\phi, \Delta\lambda)\, m(\phi', \lambda') \cos\phi'\, d\phi'\, d\lambda' = \\ = C_0 \otimes [m \cos\phi] = \mathcal{F}^{-1}\Big\{\mathcal{F}\{C_0\} \cdot \mathcal{F}\{m \cos\phi\}\Big\},$$

$$\iint C_\phi(\Delta\phi, \Delta\lambda)\, m(\phi', \lambda') \cos\phi'\, d\phi'\, d\lambda' = \\ = C_\phi \otimes [m \cos\phi] = \mathcal{F}^{-1}\Big\{\mathcal{F}\{C_\phi\} \cdot \mathcal{F}\{m \cos\phi\}\Big\},$$

requires $N^2 \times {}^2\log(N^2) = 65\,536 \times 16 =$ more than a million "stan-
8 dard operations",[8] multiplication with the coefficients $(\phi - \phi_0)$ and adding together, again 65 536 standard operations.

The grid matrices corresponding to kernel functions C_0 and C_ϕ are obtained as follows: for three reference latitudes $\phi_{-1}, \phi_0, \phi_{+1}$ we compute numerically the grids

$$C_{-1} = C(\Delta\phi, \Delta\lambda; \phi_{-1}),$$
$$C_0 = C(\Delta\phi, \Delta\lambda; \phi_0),$$
$$C_{+1} = C(\Delta\phi, \Delta\lambda; \phi_{+1}),$$

in which C_0 is directly available, and

$$C_\phi \approx \frac{C_{+1} - C_{-1}}{\phi_{+1} - \phi_{-1}}.$$

Inversion calculation is thus also directly feasible. Let ℓ be given in suitable point grid form. We compute the first approximation to m as
9 follows:[9]

[7] Fourier transforms are multiplied by multiplying the corresponding elements, see section 9.2 equation 9.3.

[8] A standard operation is a multiplication plus either an addition or a subtraction.

$$\mathcal{F}\{C_0\} \cdot \mathcal{F}\{m \cos\phi\} = \mathcal{F}\{\ell\} \quad \Longrightarrow \quad [m \cos\phi]^{(0)} = \mathcal{F}^{-1}\left\{\frac{\mathcal{F}\{\ell\}}{\mathcal{F}\{C_0\}}\right\}.$$

The second approximation is obtained by first calculating

$$\ell^{(0)} = C_0 \otimes [m \cos\phi]^{(0)} + (\phi - \phi_0) \cdot C_\phi \otimes [m \cos\phi]^{(0)},$$

after which we make the improvement

$$[m \cos\phi]^{(1)} = [m \cos\phi]^{(0)} + \mathcal{F}^{-1}\left\{\frac{\mathcal{F}\{\ell - \ell^{(0)}\}}{\mathcal{F}\{C_0\}}\right\},$$

and so on, iteratively. Two, three iterations are enough. This method has been used to compute underground mass points from gravity anomalies to represent the exterior gravity field of the Earth.[10] More is explained in Forsberg and Vermeer (1992); Vermeer (1992). [10]

9.3.4 "1-D FFT"

This is a limiting case of the previous ones, in which FFT is used only in the longitude direction. In other words, this is a zones method in which the zones have a width of only a single grid row. This method is exact if all longitudes $0° \leqslant \lambda < 360°$ are along in the calculation. It requires somewhat more computing time compared to the previous methods. In fact, it is identical to a Fourier transform in variable λ, longitude. Details are found in Haagmans et al. (1993).

9.4 Bordering and tapering of the data area

The discrete Fourier transform presupposes the data to be *periodically continuous*. In other words, it is assumed that when connecting the

[9] A Fourier transform is divided by another one by dividing the corresponding elements, see section 9.2.

[10] Because the relationship between the mass points and the observed gravity anomalies on the Earth's surface can be described exactly in geodetic co-ordinates, the method may be used with geodetic latitude φ instead of geocentric latitude ϕ. Thus, errors caused by ignoring the Earth's flattening are avoided.

eastern edge of the data area to the western edge, and the northern edge to the southern edge, the data has to be continuous across these edges.[11] In practice, this is not the case. We are faced with two different requirements:

- The data on the other side of an edge must be so far away as to have no noticeable influence across the edge on the result of the calculation.
- The data must be continuous across the edges.

Therefore, always when using FFT with the convolution theorem, two measures need to be taken.

1. We continue the data by adding a border area to the data area, so-called *bordering*. Often, the width of the border area is 25 % of the size of the data area, making the surface area of the whole calculation area four times that of the data area itself. The border is filled with measured values where those exist, otherwise with predicted (inter- or extrapolated) values.

   The calculation area for the kernel function is also made similarly four times larger. In this case the whole grid including the border area is filled with real (computed) values.

   The grid of the kernel function must be filled in such a way, that index values $i, j > N/2$ are interpreted as negative values $i - N$ and $j - N$, representing also negative Δx_i and Δy_j. Then, the peak of the function will be in the centre of the grid. If the function is symmetric, the four quadrants of the grid will look like mirror images of each other. Then, the grid will automatically be periodically continuous.
2. Because the discrete Fourier transform assumes periodicity, one must make sure that the data is continuous across the edges.

[11] Topologically the area with the edges thus connected is equivalent to a *torus*, and the data is presupposed to be continuous on the surface of the torus.

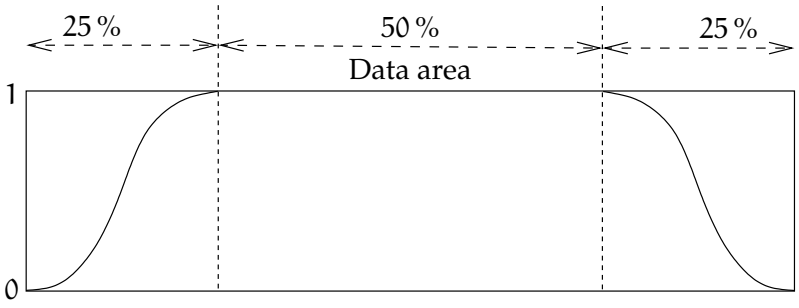


FIGURE 9.3. "Tapering" 25 %.

If the values at the edges are not zero, they may be forced to zero by multiplying the whole calculation area by a so-called *tapering* function, which goes smoothly to zero towards the edges. Such a function can easily be built: examples are a cubic spline polynomial or a Tukey or cosine taper. See figure 9.3, showing a 25 % tapering function, as well as example images 9.4, which show how non-continuity — sharply differing left and right, and upper and lower, edges — causes horizontal and vertical artefacts in the Fourier transform. These artefacts are related to the *Gibbs phenomenon*, already mentioned in section 8.9: a sharp cut-off or edge in the space domain will generate signal on all frequencies, up to the highest ones.

Gibbsin ilmiö

Many journal articles have appeared on these technicalities. Groups that were already involved in early development of FFT geoid determination in the 1980s include Forsberg's group in Copenhagen, The group of Klaus-Peter Schwarz and Michael Sideris in Calgary, Canada, the Delft group (Strang van Hees, Haagmans, De Min, Van Gelderen), the Milanese group (Sansò, Barzaghi, Brovelli), Heiner Denker at the Hannover "*Institut für Erdmessung*", and many others.

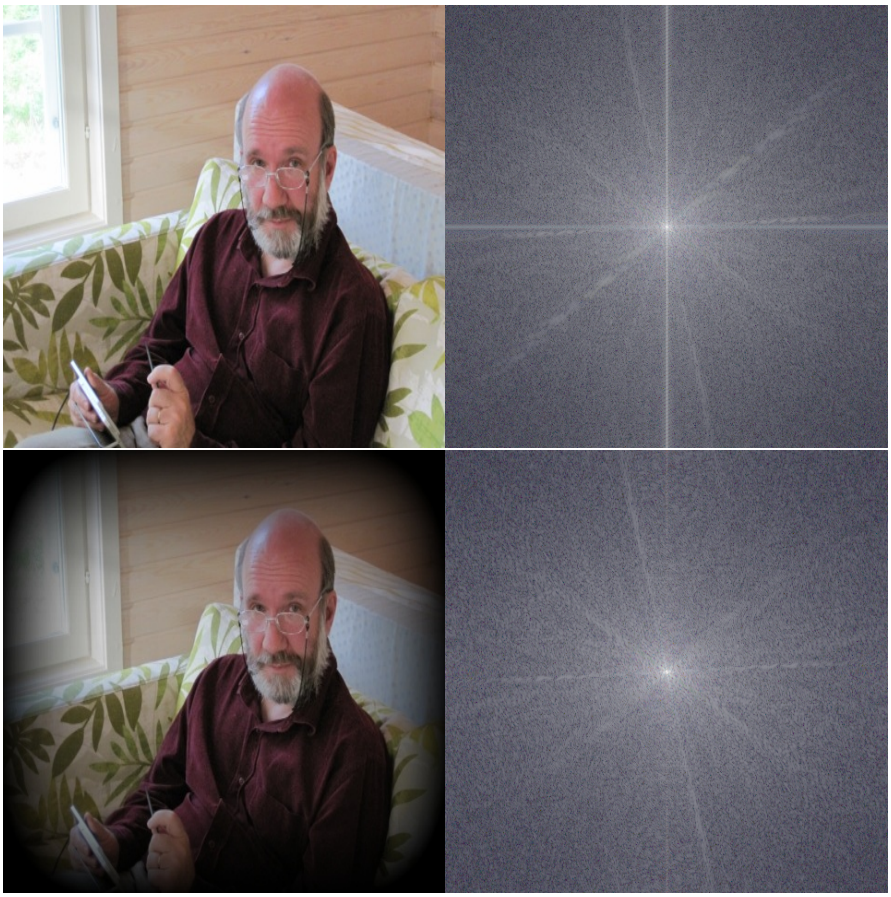


FIGURE 9.4. Example images for the FFT transform without (above) and with (below) tapering. The online FFT service from Watts (2004) was used. The images are greytone amplitude spectra $|\mathcal{F}_{uv}|$ plotted with the origin $u = v = 0$ in the centre, see appendix C.

9.5 Computing a geoid model with FFT

Nowadays computing a geoid or quasi-geoid model is easy thanks to increased computing power, especially using FFT. On the other hand, the spread of precise geodetic satellite positioning has made the availability of precise geoid models an important issue, so that one can use GNSS technology for rapid and affordable height determination.

9.5.1 GRAVSOFT software

The GRAVSOFT geoid determination software has been mainly produced in Denmark. Authors include Carl Christian Tscherning,[12] René Forsberg, Per Knudsen, the Norwegian Dag Solheim, and the Greek Dimitris Arabelos. The manual for the software is Forsberg and Tscherning (2008).

[12]

This package is in widespread use and also provides, in addition to variants of FFT geoid determination, for example least-squares collocation, as well as routines for evaluating various terrain effects. Its popularity can be partly explained by it being free for scientific use, and being distributed as source code. It is also well-documented. Therefore it has also found commercial use, for example in the petroleum extraction industry.

pienimmän neliösumman kollokaatio

GRAVSOFT has also been used a great deal for teaching, for example at many research schools organised by the IAG (International Association of Geodesy) in various countries. ISG, Geoid Schools.

9.5.2 The Finnish FIN2000 geoid

Currently two geoid models are in use in Finland: FIN2000 (figure 9.5) and FIN2005N00 (Bilker-Koivula and Ollikainen, 2009). The first model is a reference surface for the N60 height system: using it together with GNSS positioning allows determination of the N60 heights of points. The model gives geoid heights above the GRS80 reference ellipsoid. The second model is similarly a reference surface for the new N2000 height system. It, too, gives heights from the GRS80 reference ellipsoid.

vertauspinta

vertaus-ellipsoidi

The precisions (mean errors) of FIN2000 and FIN2005N00 are on the level of $\pm$ 2–3 cm.

[12] Carl Christian Tscherning (1942–2014) was a Danish physical geodesist well-known for his research into the gravity field of the Earth. He did ground-breaking work on statistical computation methods for modelling the Earth's gravity field from many different measurement types.

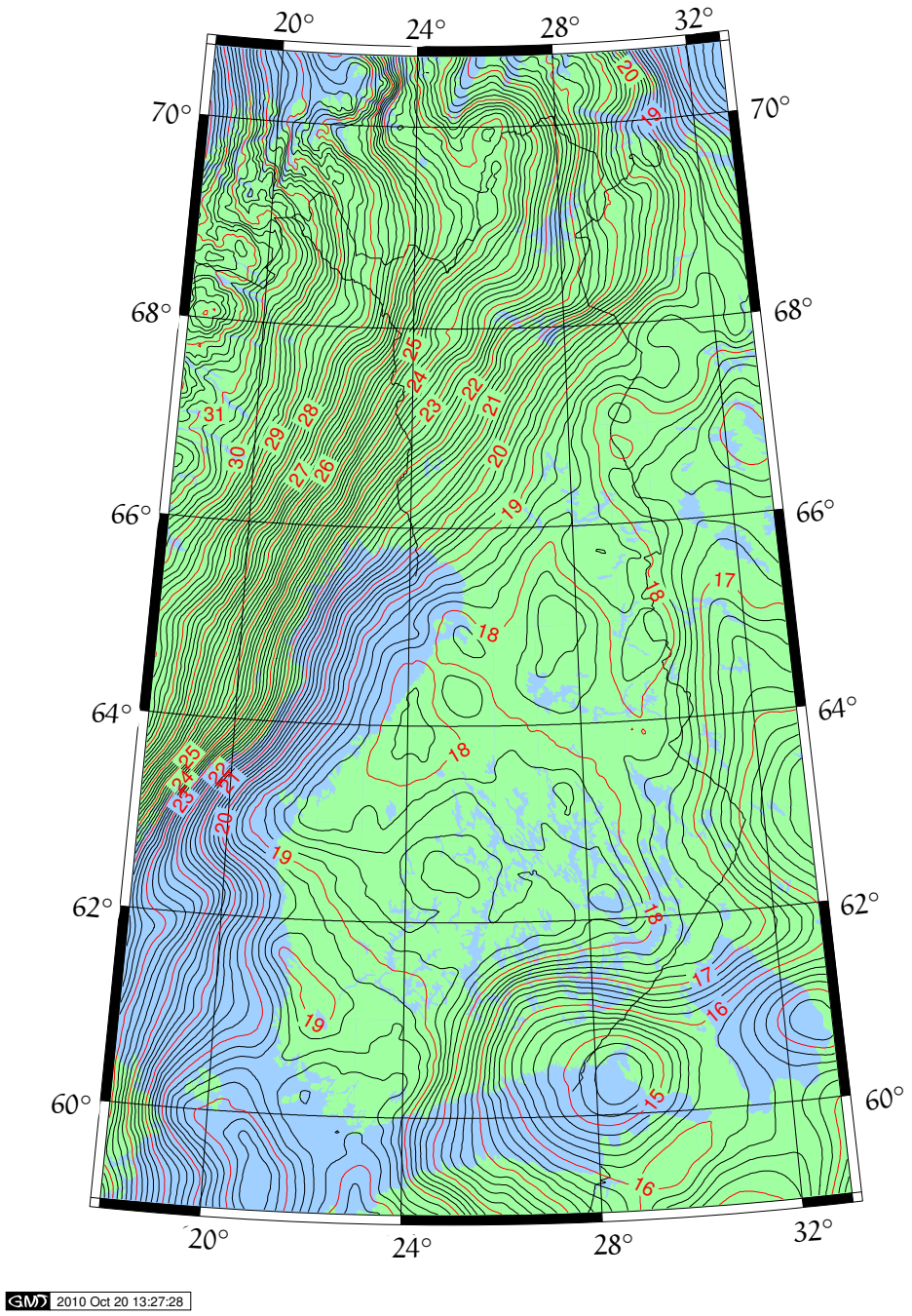


FIGURE 9.5. The Finnish FIN2000 geoid. Data © Finnish Geodetic Institute.

9.6 Use of FFT computation in other contexts

9.6.1 Satellite altimetry

The Danish researchers Per Knudsen and Ole Balthasar Andersen have computed a gravity map of the world ocean by inverting satellite altimetry derived "geoid heights" to gravity anomalies (Andersen et al., 2010). A pioneer of this method has been David Sandwell from the Scripps Institute of Oceanography in California, for example Garcia et al. (2014). The short-wavelength features in the map can tell us about the sea-floor topography.

9.6.2 Satellite gravity missions and airborne gravimetry

The data from satellite gravity missions (like CHAMP, GRACE and GOCE) can also be regionally processed using the FFT method: in the case of GOCE, the inversion of gradiometric measurements yields geoid heights on the Earth's surface from measurements made at satellite level. Airborne gravity measurements are also processed in this way using FFT. The problem is called "harmonic downwards continuation" and is in principle unstable.

ilma-gravimetria

Airborne gravimetry is a practical method for the gravimetric mapping of large areas. In the pioneering days, the gravity field over Greenland was mapped, as well as many areas around the Arctic and Antarctic. Later, areas were measured like the Brazilian Amazonas, Mongolia, and Ethiopia (Bedada, 2010), for which no full-coverage terrestrial gravimetric data existed. The strength of airborne gravimetry is that one rapidly measures large areas in a homogeneous way.

9.7 Computing terrain corrections with FFT

The *terrain correction* is a very localised phenomenon, the calculation of which requires high-resolution terrain data from a relatively small

area surrounding the computation point. Thus, calculating the terrain correction is ideally suited for the FFT method.

We show how, with FFT, we can simply and efficiently evaluate the terrain correction. We make the following simplifying *assumptions*:

- Terrain slopes are relatively gentle.
- The density ρ of the Earth's crust is constant.
- The Earth is flat — the "shoebox world".

These assumptions are not mandatory. The general case, however, leads us into a jungle of equations without aiding the conceptual picture.

The *terrain correction*, the removal of the joint effect of all the topographic masses, or lacking topographic masses, above and below the height level H of the evaluation point, can be calculated under these assumptions using the following rectangular equation, which yields the vertical component of the attraction of rock columns (figure 6.5):

$$\begin{aligned} \mathrm{TC}(x,y) &= \iint_{-\infty}^{+\infty} \frac{G\rho\left(H'(x',y') - H(x,y)\right)}{\ell^2} \cos\theta \, dx' \, dy' = \\ &= \iint_{-\infty}^{+\infty} \frac{G\rho\,(H'-H)}{\ell^2} \cdot \tfrac{1}{2}\frac{H'-H}{\ell} dx' \, dy' = \\ &= \tfrac{1}{2} G\rho \iint_{-\infty}^{+\infty} \frac{(H'-H)^2}{\ell^3} dx' \, dy'. \end{aligned} \tag{9.8}$$

Here, $G\rho\,(H'-H)/\ell^2$ is the attraction of the column and $\frac{1}{2}(H'-H)/\ell$ is the cosine of the angle θ between the force vector — assumed coming from the midpoint of the rock column — and the vertical direction. This is the so-called *prism method*.

vinoetäisyys

We will make a linear approximation, wherein ℓ, the slant distance between the evaluation point (x,y) and the moving data point (x',y'), is the horizontal distance as well:

$$\ell^2 \approx (x-x')^2 + (y-y')^2 .$$

Equation 9.8 follows straight from Newton's law of gravitation. When it is assumed that the terrain is relatively free of steep slopes, then ℓ is large compared to $H'-H$.

From equation 9.8 we obtain by expansion into terms:

$$\mathrm{TC}(x,y) = \tfrac{1}{2}G\rho H^2 \iint_{-\infty}^{+\infty} \frac{1}{\ell^3}\,dx'\,dy' - G\rho H \iint_{-\infty}^{+\infty} \frac{H'}{\ell^3}\,dx'\,dy' + \\ + \tfrac{1}{2}G\rho \iint_{-\infty}^{+\infty} \frac{(H')^2}{\ell^3}\,dx'\,dy', \quad (9.9)$$

in which every integral is a *convolution* with kernel ℓ^{-3}, and the functions to be integrated are 1, H', and $(H')^2$.

Unfortunately the function ℓ^{-3} as implicitly defined above has no Fourier transform. Therefore, we change the above definition a tiny bit by adding a small term:

$$\ell^2 = (x-x')^2 + (y-y')^2 + \delta^2. \quad (9.10)$$

The terms in the above equation 9.9 are large numbers that almost cancel each other, giving a nearly correct result. Numerically this is an unpleasant situation. There is a solution for this which we present next.

If ℓ is defined according to equation 9.10, then the Fourier transform of kernel ℓ^{-3} is (Harrison and Dickinson, 1989; Forsberg, 1984):

$$\mathcal{F}\{\ell^{-3}\} = \frac{2\pi}{\delta}\exp(-2\pi\delta q) = \frac{2\pi}{\delta}\left(1 - 2\pi\delta q + \frac{4\pi^2\delta^2 q^2}{1\cdot 2} - \cdots\right),$$

in which $q \overset{\text{def}}{=} \sqrt{\tilde{\nu}_x^2 + \tilde{\nu}_y^2} = \sqrt{u^2+v^2}\,/L$, u and v are wave indices, and $\tilde{\nu}_x = u/L$ and $\tilde{\nu}_y = v/L$ are (linear) "spatial frequencies" or wave numbers in the x and y directions in the (x,y) plane. If we substitute this into equation 9.9, we notice that the terms containing $1/\delta$ sum to zero, and of course the terms containing positive powers of δ vanish as well when $\delta \to 0$. We obtain (Harrison and Dickinson, 1989):

$$\mathcal{F}\{\mathrm{TC}\} \approx \tfrac{1}{2}G\rho H^2\,\mathcal{F}\{1\}\cdot\left(\frac{2\pi}{\delta}(1-2\pi\delta q)\right) - \\ - G\rho H\,\mathcal{F}\{H'\}\cdot\left(\frac{2\pi}{\delta}(1-2\pi\delta q)\right) + \\ + \tfrac{1}{2}G\rho\,\mathcal{F}\{(H')^2\}\cdot\left(\frac{2\pi}{\delta}(1-2\pi\delta q)\right),$$

where we left off all terms in higher powers of δ.

Re-order the terms:

$$\mathcal{F}\{\mathrm{TC}\} = \frac{2\pi}{\delta} G\rho \left(\tfrac{1}{2} H^2 \mathcal{F}\{1\} - H\mathcal{F}\{H'\} + \tfrac{1}{2}\mathcal{F}\{(H')^2\} \right) - \\ - 2\pi G\rho \cdot 2\pi q \cdot \left(\tfrac{1}{2} H^2 \mathcal{F}\{1\} - H\mathcal{F}\{H'\} + \tfrac{1}{2}\mathcal{F}\{(H')^2\} \right).$$

Because $\mathcal{F}\{1\} = 0$ if $q \neq 0$, the first term inside the second term will always vanish. We obtain (remember that H is a constant, the height of the evaluation point):

$$\mathcal{F}\{\mathrm{TC}\} = \frac{2\pi}{\delta} G\rho \left(\tfrac{1}{2}\mathcal{F}\{H^2 - HH' + \tfrac{1}{2}(H')^2\} \right) + \\ + 2\pi G\rho \cdot 2\pi q \cdot \left(H\mathcal{F}\{H'\} - \tfrac{1}{2}\mathcal{F}\{(H')^2\} \right)$$

and the reverse Fourier transform yields

$$\mathrm{TC} = \frac{\pi}{\delta} G\rho \left(H^2 - 2H'H + (H')^2 \right) + \\ + 2\pi G\rho\, \mathcal{F}^{-1}\left\{ 2\pi q \cdot \left(H\mathcal{F}\{H'\} - \tfrac{1}{2}\mathcal{F}\{(H')^2\} \right) \right\}.$$

In the first term

$$H^2 - 2H'H + (H')^2 = (H - H')^2 = 0$$

in point (x, y) in which $H' = H$, and we obtain

$$\mathrm{TC} = 4\pi^2 G\rho\, \mathcal{F}^{-1}\left\{ q \cdot \left(H\mathcal{F}\{H'\} - \tfrac{1}{2}\mathcal{F}\{(H')^2\} \right) \right\},$$

from which the troublesome $1/\delta$ has now vanished.

A *condition* for this "regularisation" or "renormalisation" is that $H' = H$ at point (x, y): the evaluation happens at the Earth's surface. The Fourier transforms above are evaluated by the FFT method.

For calculating the terrain correction TC in the *exterior space* — examples are airborne gravimetry, the effect of the sea floor at the sea surface, and the effect of the Mohorovičić discontinuity at the Earth's surface — there are techniques that express TC as a sum of convolutions, a Taylor series expansion. An early paper on this is Parker (1972).

Self-test questions

1. What is the definition of a convolution?
2. Explain the convolution theorem.
3. Check that the dimensions of the quantities on both sides of equation 9.2 match.
4. What is spatial frequency? What is the difference between linear and angular spatial frequency?
5. Explain the basic idea of the Strang van Hees method.
6. What other approaches are there to applying the FFT method on a curved (spherical or ellipsoidal) surface?
7. Why are bordering of the data area and tapering of the calculation area necessary?
8. In addition to geoid determination, where in physical geodesy is the FFT method also used?
9. When computing the terrain correction on the Earth's surface, explain the "δ trick" used in the derivation. Why is it necessary, and how does one make the δ vanish again?

Statistical methods

10

10.1 The role of uncertainty in geophysics

In geophysics, we often obtain results based on uncertain, incomplete, or otherwise deficient observational data. This also applies in the study of the Earth's gravity field: the density of gravity observations on the Earth's surface, for example, varies greatly, and large areas of the oceans and polar regions are covered only by a very sparse network of measurements. We speak of *spatial undersampling*.

Measurement technologies that work from space, on the other hand, usually provide coverage of the whole globe, oceans, poles and all. They, however, do not always measure at a very high *resolution*. Either the resolution of the method is limited — this holds for example for the gravity-field parameters calculated from satellite orbit perturbations — or the instruments measure only directly underneath the satellite's path, like satellite altimetry.

ratahäiriöt

Another often relevant uncertainty factor is that one can do precise measurements on the Earth's surface, but inside the Earth the uncertainty is much larger and the data is obtained much more indirectly.

In previous chapters we described techniques by which we could calculate desired values or parameters for the Earth's gravity field, assuming that, for example, gravity anomalies are available everywhere on the Earth's surface, and with arbitrarily high resolution. In this

chapter we look at mathematical means to handle real-world situations where this is not the case.

10.2 Linear functionals

In mathematics, a mapping that associates with every function in a given function space a certain numerical value is called a *functional*. One such is, for example, a (partial) derivative at a certain point x_0:

$$f \mapsto \frac{d}{dx} f(x)\Big|_{x=x_0}.$$

A trivial functional is also the *evaluation functional*, the function value itself (one could say the "zeroth derivative") for a certain argument value,

$$f \mapsto f(x_0).$$

Other functionals are for example the integral over a given area σ:

$$f \mapsto \int_\sigma f(x)\, dx,$$

and so on.

We may write symbolically

$$L = \frac{d}{dx}\Big|_{x=x_0}, \quad \text{meaning} \quad L\{f\} = \frac{d}{dx} f(x)\Big|_{x=x_0}.$$

A functional or operator is *linear* if

$$L\{\alpha f + \beta g\} = \alpha L\{f\} + \beta L\{g\}, \qquad \alpha, \beta \in \mathbb{R}.$$

Remember that all partial derivatives, as also the *Laplace operator* Δ, are linear.

häiriö-potentiaali

In physical geodesy, all interesting functionals are functionals of the function $T(\phi, \lambda, R) = T(\phi, \lambda, r)|_{r=R}$, that is, of the disturbing potential at the surface of a spherical Earth. The theory thus uses the spherical approximation,[1] and the surface of the sphere of radius R corresponds

[1]

to mean sea level. For example, the disturbing potential $T_P \stackrel{\text{def}}{=} T(\phi, \lambda, R)$ at a point P at sea-level location (ϕ, λ) is such a functional:

$$T(\cdot, \cdot, R) \mapsto T(\phi, \lambda, R).$$

If point P is not at sea level, a suitable functional also exists:

$$T(\cdot, \cdot, R) \mapsto T(\phi, \lambda, r).$$

If the quantity is not the disturbing potential, but, say, the gravity anomaly or the deflection of the plumb line: luotiviivan poikkeama

$$T(\cdot, \cdot, R) \mapsto \Delta g(\phi, \lambda, r), \qquad \begin{aligned} T(\cdot, \cdot, R) &\mapsto \xi(\phi, \lambda, r), \\ T(\cdot, \cdot, R) &\mapsto \eta(\phi, \lambda, r). \end{aligned}$$

All these are also *linear* functionals. In fact, if we write

$$T(\phi, \lambda, r) = \sum_{n=2}^{\infty} \frac{1}{r^{n+1}} \sum_{m=0}^{n} P_{nm}(\sin\phi) \left(a_{nm} \cos m\lambda + b_{nm} \sin m\lambda \right),$$

even the spherical-harmonic coefficients a_{nm}, b_{nm} are all linear functionals of the disturbing potential T: pallofunktiokerroin

$$T(\cdot, \cdot, R) \mapsto a_{nm}, \qquad T(\cdot, \cdot, R) \mapsto b_{nm}.$$

Here, $T(\cdot, \cdot, R)$ is shorthand for the whole function

$$T(\phi, \lambda, R), \qquad \phi \in \left[-\pi/2 \, , + \pi/2 \right], \quad \lambda \in \left[0, 2\pi \right).$$

10.3 Statistics on the Earth's surface

In statistics, we define a *stochastic process* as a stochastic quantity, or random variable, the value set or *codomain* of which is a function space. arvojoukko In other words, it is a random variable, the realisation values of which are *functions*. A stochastic process may be a quantity developing over time, the precise behaviour of which is uncertain, for example a satellite

orbit. In the same way as for a (real-valued) stochastic quantity $\underline{x}$ we
odotusarvo may define an expected value or *expectancy* $E\{\underline{x}\}$ and a *variance*

$$\Sigma_{xx} \stackrel{\text{def}}{=} \mathrm{Var}\{\underline{x}\} = E\left\{\left(\underline{x} - E\{\underline{x}\}\right)^2\right\},$$

we may also do so for a stochastic process. The only difference is that by doing so we obtain *functions*.

Let, for example, the stochastic process $\underline{x}(t)$ be a function of time. Then we may define its variance function as follows:

$$C_{xx}(t) \stackrel{\text{def}}{=} \mathrm{Var}\{\underline{x}(t)\}.$$

However, much more may be defined for a stochastic process, for example the covariance of values of the same process taken at different points in time, the *autocovariance*:

$$A_x(t_1, t_2) = C_{xx}(t_1, t_2) \stackrel{\text{def}}{=} \mathrm{Cov}\{\underline{x}(t_1), \underline{x}(t_2)\} = \\ = E\left\{\left(\underline{x}(t_1) - E\{\underline{x}(t_1)\}\right)\left(\underline{x}(t_2) - E\{\underline{x}(t_2)\}\right)\right\}.$$

Similarly if we have two different processes, we may define the *cross-covariance* between them:

$$C_{xy}(t_1, t_2) \stackrel{\text{def}}{=} \mathrm{Cov}\{\underline{x}(t_1), \underline{y}(t_2)\} = \\ = E\left\{\left(\underline{x}(t_1) - E\{\underline{x}(t_1)\}\right)\left(\underline{y}(t_2) - E\{\underline{y}(t_2)\}\right)\right\}.$$

The argument of a stochastic process is commonly *time* t. However in geophysics we study stochastic processes the arguments of which are *locations* on the Earth's surface: we talk of processes of the form $\underline{x}(\phi, \lambda)$. The definitions of auto- and cross-covariances work otherwise in the same way, but in the case of the Earth, a special problem arises. A stochastic quantity is generally defined as a quantity $\underline{x}$ from which

[1] This is not mandatory, but the error of approximation is usually small.

realisations x_1, x_2, x_3, ... are obtained, which together have certain statistical properties.

The classical example is the *dice throw*. A die can be thrown again and again, and one can practise the art of statistics on the results of the throws. Another classic example is *measurement*. Measurement of the same quantity can be repeated, and is repeated, in order to improve precision.

For a stochastic process defined on the Earth's surface, the situation is different.

We have only one Earth.

For this reason, statistics must be done in a somewhat different fashion.

Given a stochastic process — for example some geophysical quantity — on the surface of the Earth, $\underline{x}(\phi, \lambda)$, we define a quantity similar to the statistical expectancy $E\{\cdot\}$, the *geographic mean* odotusarvo

$$M\{x\} \stackrel{\text{def}}{=} \frac{1}{4\pi} \iint_\sigma x(\phi, \lambda)\, d\sigma = \frac{1}{4\pi} \int_0^{2\pi} \int_{-\pi/2}^{+\pi/2} x(\phi, \lambda) \cos\phi \, d\phi \, d\lambda. \tag{10.1}$$

Here, $x(\phi, \lambda)$ is the one and only realisation of process $\underline{x}$ that is available on this Earth.

Clearly this definition makes sense only in the case where the statistical behaviour of the process $\underline{x}(\phi, \lambda)$ is the same everywhere on Earth, independently of location (ϕ, λ). This is called the *assumption of homogeneity*. It is in fact the assumption that the spherical symmetry of the Earth extends to the statistical behaviour of field $\underline{x}$.

Similarly to the statistical variance based on expectancy, we may define the *geographic variance*:

$$C_{xx}(\phi, \lambda) = \mathrm{Var}\{x(\phi, \lambda)\} \stackrel{\text{def}}{=} M\left\{ \left(x - M\{x\} \right)^2 \right\}. \tag{10.2}$$

The global average of gravity anomalies $\Delta g(\phi, \lambda)$ *vanishes*[2] based on [2]

their definition:

$$M\{\Delta g\} = 0.$$

In that case, equation 10.2 is simplified as follows:

$$C_{\Delta g \Delta g}(\phi, \lambda) = \mathrm{Var}\{\Delta g(\phi, \lambda)\} = M\{\Delta g^2\} = \frac{1}{4\pi} \iint_\sigma \left(\Delta g(\phi, \lambda)\right)^2 \, d\sigma.$$

The definition given here of the geographic mean $M\{\cdot\}$ is based on *integration* of the one and only realisation over the surface of the Earth. As has been seen, in statistics the mean is defined slightly differently, as the *expectancy* of a stochastic process. For gravity anomalies it is $E\{\underline{\Delta g}\}$, in which $\underline{\Delta g}$ is the anomaly *considered as a stochastic process*, the series of values of Δg that results if we look at an endless series of randomly formed Earths. Not very practical!

If the expectancy of a stochastic process is the same as the mean of one realisation computed by integration — and other statistical properties are similarly the same — we speak of an *ergodic process*. Establishing empirically in geophysics that a process is ergodic is typically difficult to impossible.

10.4 The covariance function of the gravity field

Defining a *covariance function* between points P and Q is more complicated. Something like equations 10.1 and 10.2 cannot be used directly, because *both* Δg_P *and* Δg_Q:

$$\Delta g_P = \Delta g(\phi_P, \lambda_P), \qquad \Delta g_Q = \Delta g(\phi_Q, \lambda_Q),$$

can move independently over the whole Earth's surface.

In the following we assume that the covariance to be calculated will only depend on the *relative* location of points P and Q. In a homogeneous

[2] This is not exactly valid if, for example, the normal gravity field used in calculating the anomalies contains the mass of the atmosphere, but gravity values measured close to sea level do not contain the attraction of the atmosphere.

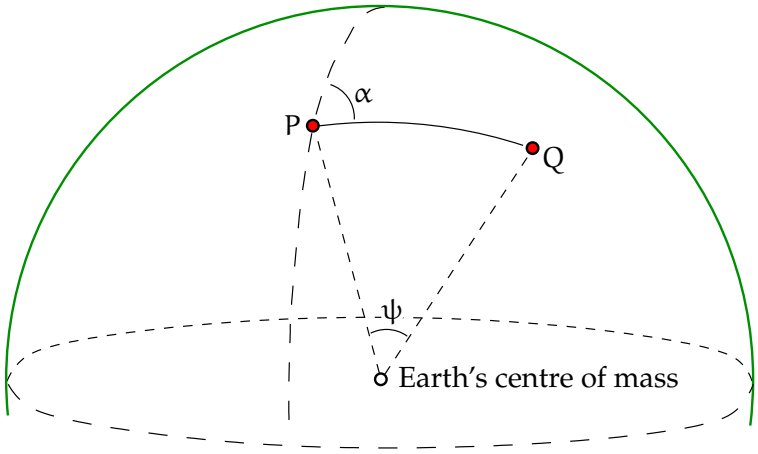


FIGURE 10.1. Definition of geocentric angular distance and azimuth.

gravity field, the covariance function will not depend on the absolute location of the points, but only on the *difference* in location between points P and Q.

Write

$$\phi_Q = \phi_Q(\phi_P, \lambda_P, \psi_{PQ}, \alpha_{PQ}), \quad \lambda_Q = \lambda_Q(\phi_P, \lambda_P, \psi_{PQ}, \alpha_{PQ}).$$

ϕ_Q and λ_Q can be computed[3] if we know ϕ_P and λ_P as well as both the *geocentric angular distance* ψ_{PQ} and the *azimuth angle* α_{PQ}. See figure 10.1.

Now we may write

$$\Delta g_Q = \Delta g_Q\Big(\phi_Q(\phi_P, \lambda_P, \psi_{PQ}, \alpha_{PQ}), \lambda_Q(\phi_P, \lambda_P, \psi_{PQ}, \alpha_{PQ})\Big) = \\ = \Delta g_Q(\phi_P, \lambda_P, \psi_{PQ}, \alpha_{PQ}),$$

and we may define as the *covariance function*

$$C_{\Delta g \Delta g}(\psi_{PQ}, \alpha_{PQ}) \stackrel{\text{def}}{=} M\Big\{\Delta g_P(\phi_P, \lambda_P)\, \Delta g_Q(\phi_P, \lambda_P, \psi_{PQ}, \alpha_{PQ})\Big\} = \\ = \frac{1}{4\pi} \iint_\sigma \Delta g_P(\phi_P, \lambda_P)\, \Delta g_Q(\phi_P, \lambda_P, \psi_{PQ}, \alpha_{PQ})\, d\sigma_P. \quad (10.3)$$

[3] This is called the *geodetic forward problem* on the sphere.

Also here, M is a geographic-mean operator. First, fix point Q in relation
4 to point P, while both azimuth α_{PQ} and distance ψ_{PQ} are held fixed.[4] Move point P, and with it, point Q, over the whole of the Earth's surface. Compute the corresponding integral over the unit sphere σ, and divide by 4π:

$$C_{\Delta g\Delta g}(\psi_{PQ}, \alpha_{PQ}) = M\{\Delta g_P \, \Delta g_{Q(P)}\} = \frac{1}{4\pi} \iint_\sigma \Delta g_P \, \Delta g_{Q(P)} \, d\sigma = \\ = \frac{1}{4\pi} \int_0^{2\pi} \int_{-\pi/2}^{+\pi/2} \Delta g_P \, \Delta g_{Q(P)} \, \cos\phi \, d\phi \, d\lambda,$$

in which $d\sigma = \cos\phi \, d\phi \, d\lambda$ is used, and $\cos\phi$ is *Jacobi's determinant* for the co-ordinates $(\phi, \lambda) = (\phi_P, \lambda_P)$ on the unit sphere.

In addition to the assumption of homogeneity, we make still the *assumption of isotropy*: the covariance function — more generally, the statistical behaviour of the gravity field — does *not* depend on the relative direction or azimuth α_{PQ} of point pair (P, Q), but only on the angular distance ψ_{PQ} between them. This, too, is, like homogeneity, one of the forms in which the Earth's spherical symmetry is expressed. In this case we may compute the geographic mean in a slightly different way, by averaging also over all azimuth angles $\alpha_{PQ} \in [0, 2\pi)$:

$$C_{\Delta g\Delta g}(\psi_{PQ}) \stackrel{\text{def}}{=} M'\{\Delta g_P \, \Delta g_{Q(P)}\} = \\ = \frac{1}{2\pi} \int_0^{2\pi} M\{\Delta g_P \, \Delta g_{Q(P)}\} \, d\alpha_{PQ} = \\ = \frac{1}{8\pi^2} \int_0^{2\pi} \int_0^{2\pi} \int_{-\pi/2}^{+\pi/2} \Delta g_P \, \Delta g_{Q(P)} \, \cos\phi \, d\phi \, d\lambda \, d\alpha_{PQ}. \quad (10.4)$$

[4] The critical reader may interject that, while the angular distance ψ_{AB} exists independently from the definition of geographical co-ordinates, that is not the case for the azimuth angle α_{AB}: it depends on the local direction of the meridian. If in equation 10.3 the angle α_{AB} is the conventional geodetic azimuth, this definition of the covariance function considers only one specific possible pattern of azimuth dependence. This suggests a generalisation to azimuth angles defined with respect to general curvilinear co-ordinates on the Earth's surface.
Also *isotropy* should then be understood as the absence of azimuth dependence not only in geographical co-ordinates, but in all possible curvilinear co-ordinates.

Remark The true gravity field of the Earth is not terribly homogeneous or isotropic, but in spite of this, both hypotheses are widely used.

10.5 Least-squares collocation

10.5.1 Stochastic processes in one dimension

pienimmän neliösumman kollokaatio

Collocation is a statistical estimation technique used to estimate the values of a *stochastic process* and calculate the uncertainties (like mean errors) of the estimates.

Let $\underline{s}(t)$ be a stochastic process, the autocovariance function of which is $C(t_i, t_j)$. Let the process furthermore be *stationary*, in other words, for any two moments in time t_i, t_j it holds that $C(t_i, t_j) = C(t_j - t_i) = C(\Delta t)$. The argument t is generally time, but could be any parameter, for example the distance travelled.

Of this process, we have *observations* made at times $t_1, t_2, \ldots, t_N$, when the corresponding process values for those times are $\underline{s}(t_1), \underline{s}(t_2), \ldots, \underline{s}(t_N)$. Let us assume, for the moment, that these values are *error-free* observations. Then the observations are function values of process $\underline{s}$, stochastic quantities, the variance matrix of which we may write as follows:

$$\mathrm{Var}\{\underline{s}_i\} = \begin{bmatrix} C(t_1,t_1) & C(t_1,t_2) & \cdots & C(t_1,t_N) \\ C(t_2,t_1) & C(t_2,t_2) & \cdots & C(t_2,t_N) \\ \vdots & \vdots & \ddots & \vdots \\ C(t_N,t_1) & C(t_N,t_2) & \cdots & C(t_N,t_N) \end{bmatrix}.$$

We also call this autocovariance matrix the *signal variance matrix* of $\underline{s}$. We use the symbol C_{ij} for this, both for one element $C_{ij} = C(t_i, t_j)$ of the matrix and for the whole matrix: $C_{ij} = \big[C(t_i, t_j), i, j = 1, \ldots, N\big]$. The symbol $\underline{s}_i$ again denotes a vector $\big[\underline{s}(t_i), i = 1, \ldots, N\big]$ consisting of process values — or one of its elements $\underline{s}(t_i)$.

Note that if the function $C(t, t')$, or $C(\Delta t)$, is known, then the whole

matrix and all of its elements can be calculated, provided all argument values (observation times) t_i are also known.

Let the shape of the problem now be that one should *estimate*, or *predict*, the value of process $\underline{s}$ at the moment in time T, that is $\underline{s}(T)$, based on our knowledge of the above-described *observations* $\underline{s}(t_i)$, $i = 1, \ldots, N$.

In the same way as we calculated above the covariances between $\underline{s}(t_i)$ and $\underline{s}(t_j)$ — elements of the signal variance matrix C_{ij} — we also compute the covariances between $\underline{s}(T)$ and all $\underline{s}(t_i)$, $i = 1, \ldots, N$. We obtain

$$\mathrm{Cov}\{\underline{s}(T), \underline{s}(t_i)\} = \begin{bmatrix} C(T,t_1) & C(T,t_2) & \cdots & C(T,t_N) \end{bmatrix}.$$

For this we may use the notation C_{Tj}. It is assumed here that there is only one point in time T for which estimation is done. Generalisation to the case where there are several T_p, $p = 1, \ldots, M$, is straightforward. In that case, the signal covariance matrix will be of size $M \times N$:

$$\mathrm{Cov}\left\{\underline{s}(T_p), \underline{s}(t_i)\right\} = \begin{bmatrix} C(T_1,t_1) & C(T_1,t_2) & \cdots & C(T_1,t_N) \\ C(T_2,t_1) & C(T_2,t_2) & \cdots & C(T_2,t_N) \\ \vdots & \vdots & & \vdots \\ C(T_M,t_1) & C(T_M,t_2) & \cdots & C(T_M,t_N) \end{bmatrix}.$$

For this we may use the more general notation C_{pj}.

10.5.2 Signal and noise

The process $\underline{s}(t)$ is called the *signal*. It is a physical phenomenon that we are *interested* in. There also exist physical phenomena that are otherwise similar, but that we are *not* interested in: on the contrary, we wish to *remove* their influence. Such stochastic processes are called *noise*.

When we make an observation, the purpose of which is to obtain a value for the quantity $\underline{s}(t_i)$, we obtain in reality a value that is not absolutely precise. The real observation thus is

$$\underline{\ell}_i = \underline{s}(t_i) + \underline{n}_i. \tag{10.5}$$

Here, $\underline{n}_i$ is a stochastic quantity: *observational error* or *noise*. Let its variance — or more precisely, the joint noise variance matrix of multiple observations — be D_{ij}. This is a similar matrix to the above C_{ij}, and also symmetric and positive definite. The only difference is that D_{ij} designates *noise*, which we are not interested in. Often, it may be assumed that the errors $\underline{n}_i$ and $\underline{n}_j$ of two different observations $\underline{\ell}_i$ and $\underline{\ell}_j$ do not correlate, in which case D_{ij} is a diagonal matrix.

10.5.3 *Estimator and variance of prediction*

Now we construct an *estimator*

$$\widehat{s}(T_p) \stackrel{\text{def}}{=} \sum_i \Lambda_{pi}\underline{\ell}_i,$$

as a linear combination of the observations $\underline{\ell}_i$ at our disposal. The purpose in life of this estimator is to get as close as possible to $\underline{s}(T_p)$. So, the quantity to be minimised is the difference

$$\widehat{s}(T_p) - \underline{s}(T_p) = \Lambda_{pi}\underline{\ell}_i - \underline{s}(T_p) = \Lambda_{pi}\left(\underline{s}(t_i) + \underline{n}_i\right) - \underline{s}(T_p).$$

Here, for the sake of writing convenience, we left the summation sign $\sum$ off (Einstein summation convention): We always sum over adjacent, identical indices, in this case i.

Study the variance of this difference, the so-called *variance of prediction*: ennustus-varianssi

$$\Sigma_{pp} \stackrel{\text{def}}{=} \mathrm{Var}\left\{\widehat{s}(T_p) - \underline{s}(T_p)\right\}.$$

We exploit *propagation of variances*, the notations introduced above, and our knowledge that surely there is no physical relationship, or *correlation*, between observation process noise $\underline{n}$ and signal $\underline{s}$: varianssien kasautuminen

$$\mathrm{Cov}\left\{\left(\underline{s}(t_i) + \underline{n}_i\right), \left(\underline{s}(t_j) + \underline{n}_j\right)\right\} = \\ = \mathrm{Cov}\left\{\underline{s}(t_i), \underline{s}(t_j)\right\} + \mathrm{Cov}\left\{\underline{n}_i, \underline{n}_j\right\} = C_{ij} + D_{ij},$$

and[5] 5

$$\Sigma_{pq} = \mathrm{Cov}\left\{\left(\widehat{s}(T_p) - \underline{s}(T_p)\right), \left(\widehat{s}(T_q) - \underline{s}(T_q)\right)\right\} =$$
$$= \Lambda_{pi}\,\mathrm{Cov}\left\{\left(\underline{s}(t_i) + \underline{n}_i\right), \left(\underline{s}(t_j) + \underline{n}_j\right)\right\}\Lambda_{jq} + \mathrm{Cov}\left\{\underline{s}(T_p), \underline{s}(T_q)\right\} -$$
$$- \Lambda_{pi}\,\mathrm{Cov}\left\{\underline{s}(t_i), \underline{s}(T_q)\right\} - \mathrm{Cov}\left\{\underline{s}(T_p), \underline{s}(t_j)\right\}\Lambda_{jq} =$$
$$= \Lambda_{pi}\left(C_{ij} + D_{ij}\right)\Lambda_{jq} + C_{pq} - \Lambda_{pi}C_{iq} - C_{pj}\Lambda_{jq}. \quad (10.6)$$

The variances, or diagonal elements, Σ_{pp} of the matrix are now obtained by setting $q = p$.

10.5.4 *Showing optimality*

Here we show that the optimal estimator is indeed the one producing the minimum possible variances. Choose

$$\Lambda_{pj} \stackrel{\text{def}}{=} C_{pi}\left(C_{ij} + D_{ij}\right)^{-1}.$$

Then, from equation 10.6 and exploiting the symmetry of the C and D matrices, we obtain

$$\Sigma_{pp} = C_{pi}\left(C_{ij} + D_{ij}\right)^{-1} C_{jp} + C_{pp} -$$
$$- C_{pi}\left(C_{ij} + D_{ij}\right)^{-1} C_{jp} - C_{pi}\left(C_{ij} + D_{ij}\right)^{-1} C_{jp} =$$
$$= C_{pp} - C_{pi}\left(C_{ij} + D_{ij}\right)^{-1} C_{jp}. \quad (10.7)$$

Let us study next the *alternative choice*

$$\Lambda_{pj} = C_{pi}\left(C_{ij} + D_{ij}\right)^{-1} + \delta\Lambda_{pj}.$$

In this case we obtain by substitution

$$\Sigma'_{pp} = \overbrace{\Lambda_{pi}\left(C_{ij} + D_{ij}\right)\Lambda_{jp}}^{\text{I}} + C_{pp}\overbrace{-\Lambda_{pj}C_{jp}}^{\text{II}}\overbrace{-C_{pi}\Lambda_{ip}}^{\text{III}},$$

in which

[5] The matrix C_{iq} is the transpose of C_{pj}, the matrix Λ_{jq} the transpose of Λ_{pi}.

$$\begin{aligned} \mathrm{I} &= \Lambda_{pi}\,(C_{ij}+D_{ij})\,\Lambda_{jp} = \\ &= \left(C_{pi}\,(C_{ij}+D_{ij})^{-1}+\delta\Lambda_{pj}\right)(C_{ij}+D_{ij})\left((C_{jk}+D_{jk})^{-1}\,C_{kp}+\delta\Lambda_{kp}\right) = \\ &= C_{pi}\,(C_{ij}+D_{ij})^{-1}\,C_{jp} + C_{pi}\delta\Lambda_{ip} + \delta\Lambda_{pi}C_{ip} + \delta\Lambda_{pi}\,(C_{ij}+D_{ij})\,\delta\Lambda_{jp}\,, \end{aligned}$$

$$\begin{aligned} \mathrm{II} &= -\Lambda_{pj}C_{jp} = -\left(C_{pi}\,(C_{ij}+D_{ij})^{-1}+\delta\Lambda_{pj}\right)C_{jp} = \\ &= -C_{pi}\,(C_{ij}+D_{ij})^{-1}\,C_{jp} - \delta\Lambda_{pi}C_{ip} \end{aligned}$$

and

$$\begin{aligned} \mathrm{III} &= -C_{pi}\Lambda_{ip} = -C_{pi}\left((C_{ij}+D_{ij})^{-1}\,C_{jp}+\delta\Lambda_{ip}\right) = \\ &= -C_{pi}\,(C_{ij}+D_{ij})^{-1}\,C_{jp} - C_{pi}\delta\Lambda_{ip}, \end{aligned}$$

with the final result

$$\Sigma'_{pp} = C_{pp} \overbrace{-C_{pi}\,(C_{ij}+D_{ij})^{-1}\,C_{jp}}^{\mathrm{III}} \overbrace{+\delta\Lambda_{pi}\,(C_{ij}+D_{ij})\,\delta\Lambda_{jp}}^{\mathrm{I}}.$$

Here, the last term — the only difference from result 10.7 — is positive, because the matrices C_{ij} and D_{ij} are positive definite: $\Sigma'_{pp} > \Sigma_{pp}$, except when $\delta\Lambda_{pi} = 0$. In other words, the solution given above,

$$\Lambda_{pj} = C_{pi}\,(C_{ij}+D_{ij})^{-1} \quad\Longrightarrow\quad \widehat{s}(T_p) = C_{pi}\,(C_{ij}+D_{ij})^{-1}\,\underline{\ell}_j,$$

is *optimal* in the sense of least-squares — more precisely, in the sense of minimising the variance of prediction Σ_{pp}.

10.5.5 The covariance function of gravity anomalies

Least-squares collocation is used extensively to optimally estimate gravity values and other functionals of the Earth's surface.

If we have two points, P and Q, with measured gravity anomalies $\underline{\Delta g}_P = \underline{\Delta g}(\phi_P, \lambda_P)$ and $\underline{\Delta g}_Q = \underline{\Delta g}(\phi_Q, \lambda_Q)$, one would like to determine the covariance between these two anomalies,

$$\mathrm{Cov}\left\{\underline{\Delta g}_P, \underline{\Delta g}_Q\right\}.$$

As argued in section 10.4, we can only empirically derive such a covariance by looking at all point pairs (P, Q) that are in the same relative position around the globe, and averaging over them using the M or M' operator.

Normally the covariance is assumed to depend *only* on the geocentric angular distance ψ between points P and Q. Then, we speak of an *isotropic process* $\underline{\Delta g}(\phi, \lambda)$. The covariance will be

$$\mathrm{Cov}\{\Delta g_P, \Delta g_Q\} = M'\{\Delta g_P \Delta g_{Q(P)}\} = C(\psi_{PQ}).$$

[6] A popular covariance function for gravity anomalies is Hirvonen's[6] covariance function:

$$C(\psi) = \frac{C_0}{1 + \left(\psi/\psi_0\right)^2}, \tag{10.8}$$

in which $C_0 = C(0)$ and ψ_0 are parameters describing the behaviour of the gravity field. $C_0 = \mathrm{Var}\{\Delta g(\phi, \lambda)\} = M\{\Delta g^2\}$ is called the *signal variance*, ψ_0 the *correlation length*. ψ_0 gives the distance at which the
[7] correlation between the gravity anomalies in two points is still 50 %.[7]

In local applications, instead of the angular distance ψ one uses the linear distance

$$s = R\,\psi,$$

where R is the radius of the Earth. Then

$$C(s) = \frac{C_0}{1 + \left(s/d\right)^2}.$$

[6] Reino Antero Hirvonen (1908–1989) was a Finnish physical and mathematical geodesist.

This equation was derived from gravimetric data for Ohio state, USA, but it has broader validity. $C(0) = C_0$, the signal variance. The variable $d = R\psi_0$ is also called the correlation length. It is the distance d for which $C(d) = \frac{1}{2}C_0$, as seen from the equation.

The quantity C_0 varies considerably between areas, from hundreds to thousands of mGal[2], and tends to be largest in mountainous areas. The quantity d is generally of the order of magnitude of tens of kilometres.

Alternative functions that are also often used in local applications include the covariance function of the stationary Gauss–Markov process, and a quadratic variant:

$$C(\psi) = C_0 \exp\left(-\frac{\psi}{\psi_0}\right), \qquad C(\psi) = C_0 \exp\left(-\left(\frac{\psi}{\psi_0}\right)^2\right).$$

10.5.6 Least-squares collocation for gravity anomalies

If N points P_i, $i = 1, \ldots, N$ are given, where were measured gravity values and calculated anomalies — $\underline{\Delta g}_i = \underline{\Delta g}(\phi_i, \lambda_i)$, we may, as above, construct a *signal variance matrix*

$$C_{ij} \stackrel{\text{def}}{=} \operatorname{Var}\left\{\underline{\Delta g}_i\right\} =$$
$$= \begin{bmatrix} C_0 & C(\psi_{12}) & \cdots & C(\psi_{1N}) \\ C(\psi_{21}) & C_0 & \cdots & C(\psi_{2N}) \\ \vdots & \vdots & & \vdots \\ C(\psi_{N1}) & C(\psi_{N2}) & \cdots & C_0 \end{bmatrix} = \begin{bmatrix} C_0 & C_{12} & \cdots & C_{1N} \\ C_{21} & C_0 & \cdots & C_{2N} \\ \vdots & \vdots & & \vdots \\ C_{N1} & C_{N2} & \cdots & C_0 \end{bmatrix},$$

in which all elements $C(\psi_{ij})$ are calculated using covariance function 10.8 given above.

---

[7] The correlation is

$$\operatorname{Corr}\left\{\underline{\Delta g}_P, \underline{\Delta g}_Q\right\} = \frac{\operatorname{Cov}\left\{\underline{\Delta g}_P, \underline{\Delta g}_Q\right\}}{\sqrt{\operatorname{Var}\left\{\underline{\Delta g}_P\right\} \operatorname{Var}\left\{\underline{\Delta g}_Q\right\}}} = \frac{\dfrac{C_0}{1 + (\psi/\psi_0)^2}}{\sqrt{C_0 C_0}} = \frac{1}{1 + (\psi/\psi_0)^2},$$

which is 0.5 for $\psi = \psi_0$.

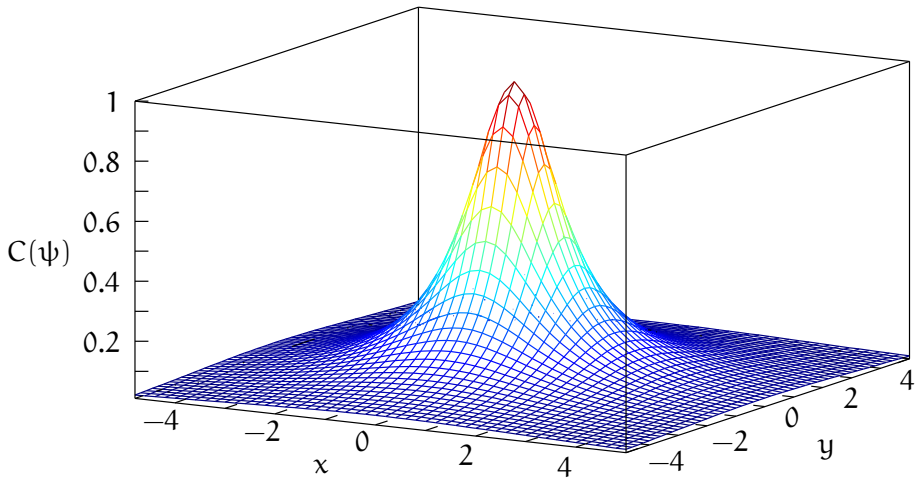


FIGURE 10.2. Hirvonen's covariance function in two dimensions. $C_0 = \psi_0 = 1$ is assumed.

If we also compute for the point P at which gravity is unknown:

$$\mathrm{Cov}\left\{\underline{\Delta g}_P, \underline{\Delta g}_i\right\} = \left[\begin{array}{cccc} C(\psi_{P1}) & C(\psi_{P2}) & \cdots & C(\psi_{PN}) \end{array}\right] \stackrel{\mathrm{def}}{=} C_{Pi},$$

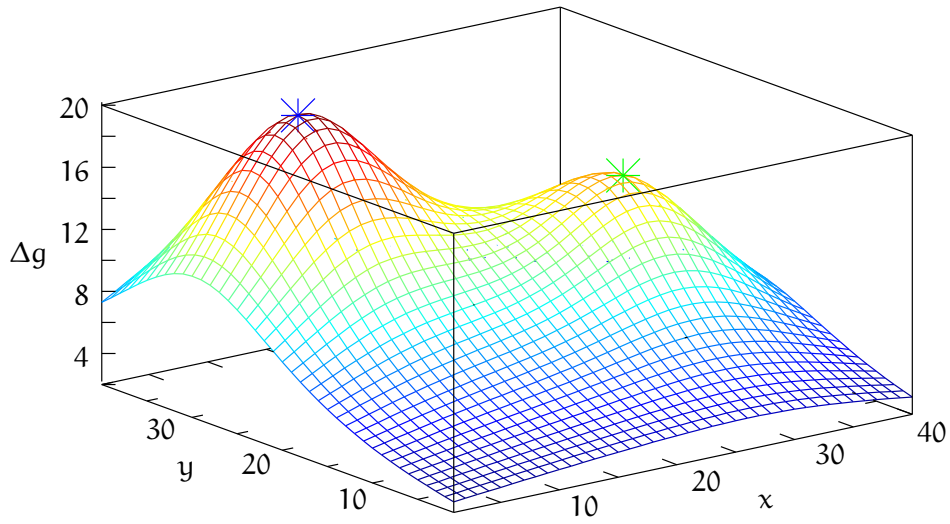


FIGURE 10.3. An example of least-squares collocation. Here are given two data points $\underline{\Delta g}_1$ and $\underline{\Delta g}_2$ (stars); the surface plotted gives the estimated value $\widehat{\Delta g}_P$ for each point P in the area. We use least-squares collocation for inter- and extrapolating gravimetric data.

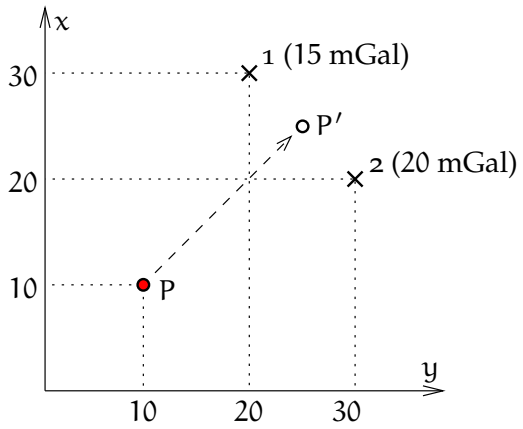


FIGURE 10.4. Least-squares collocation, calculation example.

we obtain, in the same way as before, for the *least-squares collocation* solution

$$\widehat{\Delta g}_P = C_{Pi}\,(C_{ij} + D_{ij})^{-1}\,\underline{\ell}_j \approx C_{Pi} C_{ij}^{-1} \underline{\ell}_j,$$

in which the $\underline{\ell}_j = \underline{\Delta g}_j + \underline{n}_j$ are gravity anomaly observations made in points $j = 1, \ldots, N$. The matrix D_{ij}, which we leave out of consideration, again describes the random observation error, observation uncertainty, or *noise* $\underline{n}_i$ associated with making those observations. Often D_{ij} is a diagonal matrix, meaning that the observations are statistically independent and do not correlate with each other.

We may also compute a precision assessment of this solution, the *variance of prediction*, equation 10.11:

$$\Sigma_{PP} = C_0 - C_{Pi}\,(C_{ij} + D_{ij})^{-1}\,C_{jP} \approx C_0 - C_{Pi} C_{ij}^{-1} C_{jP}$$

in the case of one unknown prediction point P. Its square root

$$\sigma_{\Delta g_P} = \sqrt{\Sigma_{PP}}$$

is the *mean error* of estimator $\widehat{\Delta g}_P$.

10.5.7 *Calculation example*

See figure 10.4. Two points are given where gravity has been measured and gravity anomalies calculated: $\underline{\Delta g}_1 = 15\,\text{mGal}$, $\underline{\Delta g}_2 = 20\,\text{mGal}$. The co-ordinates in the x and y directions are in kilometres. It is assumed that between the gravity anomalies of different points, Hirvonen's covariance function,

$$C(s) = \frac{C_0}{1 + \left(s/d\right)^2},$$

applies, in which $d = 20\,\text{km}$ and $C_0 = \pm 1000\,\text{mGal}^2$. In addition, it is assumed that the gravity measurements done — including height determination of the gravity points! — are errorless. So, $D_{ij} = 0$, i, $j = 1, 2$.

Kysymys Calculate an estimate of the gravity anomaly $\widehat{\Delta g}_P$ at point P and its mean error $\sigma_{\Delta g_P} = \sqrt{\Sigma_{PP}}$.

Vastaus Calculate first the distances s and the corresponding covariances C.

$$\begin{aligned}
s_{12}^2 &= \left((30-20)^2 + (20-30)^2\right)\,\text{km}^2 = 200\,\text{km}^2,\\
C_{12} = C_{21} &= \frac{1000\,\text{mGal}^2}{1 + 200/400} = 666.66\ldots\,\text{mGal}^2,\\
s_{1P}^2 &= \left((30-10)^2 + (20-10)^2\right)\,\text{km}^2 = 500\,\text{km}^2,\\
C_{1P} &= \frac{1000\,\text{mGal}^2}{1 + 500/400} = 444.44\ldots\,\text{mGal}^2,\\
s_{2P}^2 &= \left((20-10)^2 + (30-10)^2\right)\,\text{km}^2 = 500\,\text{km}^2,\\
C_{2P} &= \frac{1000\,\text{mGal}^2}{1 + 500/400} = 444.44\ldots\,\text{mGal}^2.
\end{aligned}$$

From this follows

$$\begin{aligned}
C_{ij} + D_{ij} &\approx\\
&\approx C_{ij} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} = \begin{bmatrix} 1000 & 666.66 \\ 666.66 & 1000 \end{bmatrix}\,\text{mGal}^2,
\end{aligned}$$

and its inverse matrix

$$(C_{ij} + D_{ij})^{-1} = \begin{bmatrix} 0.0018 & -0.0012 \\ -0.0012 & 0.0018 \end{bmatrix} \text{mGal}^{-2}.$$

Furthermore

$$C_{Pi} = \begin{bmatrix} C_{P1} & C_{P2} \end{bmatrix} = \begin{bmatrix} 444.44 & 444.44 \end{bmatrix} \text{mGal}^2.$$

As the vector of observations is

$$\underline{\Delta g}_j = \begin{bmatrix} \underline{\Delta g}_1 \\ \underline{\Delta g}_2 \end{bmatrix} = \begin{bmatrix} 15 \\ 20 \end{bmatrix} \text{mGal},$$

we obtain the result

$$\widehat{\Delta g}_P = \\ = \begin{bmatrix} 444.44 & 444.44 \end{bmatrix} \begin{bmatrix} 0.0018 & -0.0012 \\ -0.0012 & 0.0018 \end{bmatrix} \begin{bmatrix} 15 \\ 20 \end{bmatrix} \text{mGal} = \\ = 9.333\,\text{mGal}.$$

The precision, the *variance of prediction*, equation 10.11:

$$\Sigma_{PP} = C_{PP} - C_{Pi}\,(C_{ij} + D_{ij})^{-1}\,C_{jP} = \\ = C_0 - \begin{bmatrix} 444.44 & 444.44 \end{bmatrix} \begin{bmatrix} 0.0018 & -0.0012 \\ -0.0012 & 0.0018 \end{bmatrix} \begin{bmatrix} 444.44 \\ 444.44 \end{bmatrix} \text{mGal}^2 = \\ = 762.96\,\text{mGal}^2,$$

so

$$\sigma_{\Delta g_P} = \sqrt{\Sigma_{PP}} = \pm 27.622\,\text{mGal}.$$

Summarising the result:

$$\widehat{\Delta g}_P = 9.333 \pm 27.622\,\text{mGal}.$$

Observe that the gravity anomaly estimate found is much smaller than its own uncertainty, and thus *does not differ significantly from*

zero. In fact, not using the observational data at all would leave us with the *a priori* estimate

$$\widehat{\Delta g}_P = 0 \pm \sqrt{1000}\,\text{mGal} = 0 \pm 31.623\,\text{mGal},$$

almost as good.

If, instead, we had used point P′ in between points 1 and 2, at location (25 km, 25 km), then

$$C_{P'1} = C_{P'2} = 1000\,\text{mGal}^2 \Big/ \left(1 + {}^{50}/_{400}\right) = 888.89\,\text{mGal}^2$$

and $\widehat{\Delta g}_{P'} = 18.667 \pm 7.201$ mGal, which is clearly better than the *a priori* estimate of zero.

And if we had chosen instead the Gauss–Markov covariance function

$$C = C_0 \exp(-{}^{s}/_{d}),$$

we would have obtained the results $\widehat{\Delta g}_P = 7.664 \pm 29.272$ mGal for the original point location, and $\widehat{\Delta g}_{P'} = 16.460 \pm 18.426$ mGal for the shifted point location.

10.5.8 *Theory of least-squares collocation*

pienimmän neliösumman kollokaatio

Above we presented one popular application of least-squares collocation. Here we look at the method more generally. The basic equation is

$$\widehat{\mathbf{f}} = C_{fg}\left(C_{gg} + D_{gg}\right)^{-1}\left(\underline{\mathbf{g}} + \underline{\mathbf{n}}\right). \tag{10.9}$$

The vector $\underline{\mathbf{g}}$ contains *observed quantities* $\underline{g}_i$, the vector $\underline{\mathbf{n}}$ contains the observational errors: uncertainty or *noise*, and $\widehat{\mathbf{f}}$ is a vector of quantities $\widehat{f}_p$ to be *predicted*.

Both vectors $\underline{\mathbf{g}}$ and $\widehat{\mathbf{f}}$ can, for example, be gravity anomalies, in which case we have *homogeneous prediction*, a type of inter- or extrapolation. More generally $\widehat{\mathbf{f}}$ and $\underline{\mathbf{g}}$ are of different types: for example $\widehat{\mathbf{f}}$ consists of

geoid heights $\underline{N}_p$ and $\underline{\mathbf{g}}$ of gravity anomalies $\underline{\Delta g}_i$. In the latter case, the Stokes equation is "covertly" along in the structure of the C matrices.

These matrices are built from covariance functions. Their elements can be expressed as follows:[8]

$$\left[C_{fg}\right]_{pi} = M'\{f_p g_i\}, \quad \left[C_{gg}\right]_{ij} = M'\{g_i g_j\}, \quad \left[D_{gg}\right]_{ij} = E\{\underline{n}_i \underline{n}_j\},$$

in which $\underline{n}_i$, an element of vector $\underline{\mathbf{n}}$, represents the uncertainty of the observation process appearing in observation equation 10.5:

$$\underline{\ell}_i = g_i + \underline{n}_i, \quad \text{or equivalently} \quad \underline{\boldsymbol{\ell}} = \mathbf{g} + \underline{\mathbf{n}}.$$

$\underline{\boldsymbol{\ell}}$ is the vector of the observation values themselves, including observation uncertainty $\underline{\mathbf{n}}$.

The D matrix is the variance matrix of observational uncertainty, the *noise variance matrix* describing a property of the observational process, not of the gravity field. While the values of $M'\{\Delta g_i \, \Delta g_j\}$ can be as large as 1200 mGal^2, the values of $E\{\underline{n}_i \underline{n}_j\}$ can be much smaller, depending on the measurement technique used, for example as small as 0.01 mGal^2.

This does not apply however in the case of block averages — for example averages over blocks of size $1^\circ \times 1^\circ$, computed from scattered measurements — which are often very imprecise.

The great strength of least-squares collocation is its *flexibility*. Different observation types may be handled with a single unified theory and method, the locations of observation points are free, and the result is obtained directly as freely choosable quantities in locations where one wants them.

---

[8] Here, we use the *geographic mean* $M'\{\cdot\}$ for evaluating the signal covariances. In doing so, $\mathbf{f}$ and $\mathbf{g}$ are no longer considered stochastic. It is assumed that their global geographic mean vanishes: $M\{\mathbf{f}\} = M\{\mathbf{g}\} = 0$.

10.6 Prediction of gravity anomalies

If the quantity to be calculated or estimated, $\widehat{\mathbf{f}}$, is of the same type as the observed quantity, $\underline{\mathbf{g}}$, one speaks of *homogeneous prediction*. For example, the prediction equation for gravity anomalies already presented in subsection 10.5.6 is obtained from equation 10.9 by substitution:

$$\widehat{\Delta g}_P = C_{Pi} \left(C_{ij} + D_{ij}\right)^{-1} \underline{\ell}_j. \tag{10.10}$$

Here are several points j where gravity has been measured: let us say, N observations $\underline{\ell}_j = \underline{\Delta g}_j + \underline{n}_j$, $j = 1, \ldots, N$. The number of points to be predicted may be one: point P, or many. The matrices C_{ij} and D_{ij} are square, and the inverse of their sum exists. C_{Pi} is a rectangular matrix. If there is only one point P, C_{Pi} is a size $1 \times N$ row matrix.

[9] The *prediction error* is now the difference quantity[9] $\widehat{\Delta g}_P - \underline{\Delta g}_P$, and its variance ("variance of prediction") is

$$\begin{aligned}\Sigma_{PP} &\overset{\text{def}}{=} \operatorname{Var}\left\{\widehat{\Delta g}_P - \underline{\Delta g}_P\right\} = \\ &= \operatorname{Var}\left\{\widehat{\Delta g}_P\right\} + \operatorname{Var}\left\{\underline{\Delta g}_P\right\} - \operatorname{Cov}\left\{\widehat{\Delta g}_P, \underline{\Delta g}_P\right\} - \operatorname{Cov}\left\{\underline{\Delta g}_P, \widehat{\Delta g}_P\right\}.\end{aligned}$$

Here (propagation of variances applied to equation 10.10):

$$\begin{aligned}\operatorname{Var}\left\{\widehat{\Delta g}_P\right\} = C_{Pi} \left(C_{ij} + D_{ij}\right)^{-1} \left(C_{jk} + D_{jk}\right) \left(C_{k\ell} + D_{k\ell}\right)^{-1} C_{\ell P} = \\ = C_{Pi} \left(C_{ij} + D_{ij}\right)^{-1} C_{jP}\end{aligned}$$

and

$$\begin{aligned}\operatorname{Cov}\left\{\widehat{\Delta g}_P, \underline{\Delta g}_P\right\} = \operatorname{Cov}\left\{C_{Pi} \left(C_{ij} + D_{ij}\right)^{-1} \left(\underline{\Delta g}_j + \underline{n}_j\right), \underline{\Delta g}_P\right\} = \\ = C_{Pi} \left(C_{ij} + D_{ij}\right)^{-1} \left(\operatorname{Cov}\left\{\underline{\Delta g}_j, \underline{\Delta g}_P\right\} + 0\right) = \\ = C_{Pi} \left(C_{ij} + D_{ij}\right)^{-1} C_{jP},\end{aligned}$$

[9] Be aware that here, $\underline{\Delta g}_P$ is the *true value* of the gravity anomaly at point P, which we do not know empirically. The *measured value* would be $\underline{\ell}_P = \underline{\Delta g}_P + \underline{n}_P$, in which $\underline{n}_P$ is the random error or "noise" of the gravimetric observation.

and also

$$\mathrm{Cov}\left\{\underline{\Delta g}_P, \widehat{\Delta g}_P\right\} = C_{Pi}(C_{ij} + D_{ij})^{-1} C_{jP}$$

and finally, the signal variance $\mathrm{Var}\left\{\underline{\Delta g}_P\right\} = C_{PP}$.

Here, C_{iP} (also called C_{jP}, or even $C_{\ell P}$) is the transpose of C_{Pi}. The matrix $(C_{ij} + D_{ij})^{-1}$ is symmetric and its own transpose.

The end result is

$$\begin{aligned}\Sigma_{PP} = C_{Pi}(C_{ij} + D_{ij})^{-1} C_{jP} + C_{PP} - \\ - C_{Pi}(C_{ij} + D_{ij})^{-1} C_{jP} - C_{Pi}(C_{ij} + D_{ij})^{-1} C_{jP} = \\ = C_{PP} - C_{Pi}(C_{ij} + D_{ij})^{-1} C_{jP}.\end{aligned}$$

In case $D_{ij} \ll C_{ij}$, we obtain a simpler, often-used result:

$$\Sigma_{PP} \approx C_{PP} - C_{Pi} C_{ij}^{-1} C_{jP}. \tag{10.11}$$

Borderline cases

- Point P is far from all points i. Then $C_{Pi} \approx 0$ and $\Sigma_{PP} \approx C_{PP}$, so prediction is impossible in practice, and the prediction equation 10.10 will yield the value zero. The *mean error* of prediction $\sigma_{\Delta g_P} = \sqrt{\Sigma_{PP}}$ is the same as the variability $\sqrt{C_{PP}}$ of the gravity anomaly signal, the square root of the signal variance.
- Point P is identical with one of the points i. Then, if we use only that point i, we obtain

  $$\Sigma_{PP} = C_{PP} - C_{PP} C_{PP}^{-1} C_{PP} = 0,$$

  no prediction error whatsoever — as the value at the prediction point was already known!.

  However, if $D_{PP} \neq 0$ (but small), the result is $\Sigma_{PP} \approx D_{PP}$.

10.7 Covariance function and degree variances

10.7.1 *The covariance function of the disturbing potential*

In theoretical work we use, instead of gravity anomalies, the covariance function of the *disturbing potential* T on the Earth's surface:

$$K(P,Q) = K(\psi_{PQ}, \alpha_{PQ}) \stackrel{\text{def}}{=} M\{T_P T_{Q(P)}\},$$

or alternatively using equation 10.4:

$$K(P,Q) = K(\psi_{PQ}) \stackrel{\text{def}}{=} M'\{T_P T_{Q(P)}\} = \\ = \frac{1}{8\pi^2}\int_0^{2\pi}\int_0^{2\pi}\int_{-\pi/2}^{+\pi/2} T_P T_{Q(P)} \cos\phi \, d\phi \, d\lambda \, d\alpha_{PQ}. \quad (10.12)$$

Here it is assumed that the disturbing potential is *isotropic*: K does not depend on α but only on ψ.

We choose on the unit sphere a co-ordinate system where point P is a "pole". In this system, the parameters α_{PQ} and ψ_{PQ} are the spherical co-ordinates of point Q. The covariance function is expanded into the following sum:

$$K(\psi) = \sum_{n=2}^{\infty}\sum_{m=-n}^{n} k_{nm} Y_{nm}(\psi, \alpha)$$

with Y_{nm} defined as in equation 3.3:

$$Y_{nm}(\psi,\alpha) = \begin{cases} P_{nm}(\cos\psi)\cos m\alpha & \text{if } m \geqslant 0, \\ P_{n|m|}(\cos\psi)\sin|m|\,\alpha & \text{if } m < 0. \end{cases} \quad (10.13)$$

Based on isotropy, all coefficients for which the order $m \neq 0$, vanish: the expressions on the right-hand side of equation 10.13 can only be independent of α if $m = 0$. So

$$K(\psi) = \sum_{n=2}^{\infty} k_{n0} Y_{n0}(\psi) = \sum_{n=2}^{\infty} k_n P_n(\cos\psi). \quad (10.14)$$

The coefficients k_n are called the *degree variances* (of the disturbing potential). For isotropic covariance functions $K(\psi)$, the information content of the degree variances k_n, $n = 2, 3, \ldots$ is the same as that of the function itself, and is in fact its *spectral representation*. astevarianssi

10.7.2 Degree variances and spherical-harmonic coefficients

Multiply equation 10.14 with $P_{n'}(\cos\psi)\sin\psi$ and integrate:

$$\int_0^\pi K(\psi)\,P_{n'}(\cos\psi)\sin\psi\,d\psi = \\ = \sum_{n=2}^{\infty} k_n \int_0^\pi P_n(\cos\psi)\,P_{n'}(\cos\psi)\sin\psi\,d\psi = \\ = \sum_{n=2}^{\infty} k_n \int_{-1}^{+1} P_n(t)\,P_{n'}(t)\,dt = k_{n'}\,\frac{2}{2n+1},$$

using orthogonality condition 3.5. It follows that

$$k_n = \frac{2n+1}{2}\int_0^\pi K(\psi)\,P_n(\cos\psi)\sin\psi\,d\psi, \qquad (10.15)$$

meaning that, if $K(\psi)$ is given, we can calculate all k_n.

Substituting $K(\psi_{PQ})$ from equation 10.12 yields, with abbreviations $\psi = \psi_{PQ}$, $\alpha = \alpha_{PQ}$:

$$k_n = \frac{2n+1}{2}\int_0^\pi \overbrace{\frac{1}{8\pi^2}\int_0^{2\pi}\int_0^{2\pi}\int_{-\pi/2}^{+\pi/2} T_P\,T_{Q(P)}\cos\phi\,d\phi\,d\lambda\,d\alpha}^{K(\psi)}\,P_n(\cos\psi)\sin\psi\,d\psi = \\ = \frac{2n+1}{16\pi^2}\int_0^{2\pi}\int_{-\pi/2}^{+\pi/2} T_P \overbrace{\int_0^{2\pi}\int_0^\pi T_{Q(P)}\,P_n(\cos\psi)\sin\psi\,d\psi\,d\alpha}^{I}\cos\phi\,d\phi\,d\lambda.$$

Here we have interchanged the order of the integrals, as is allowed, and moved T_P to another place.

The expression I is a surface integral over the unit sphere:

$$I = \int_0^{2\pi}\int_0^\pi T_{Q(P)}P_n(\cos\psi_{PQ})\sin\psi_{PQ}\,d\psi_{PQ}\,d\alpha_{PQ} = \\ = \iint_\sigma T_{Q(P)}\,P_n(\cos\psi_{PQ})\,d\sigma_Q = \frac{4\pi}{2n+1}T_{n,P},$$

in which $T_{n,P} = T_n(\phi_P, \lambda_P) = T_n(\phi, \lambda)$. T_n is the constituent of the disturbing potential T for the harmonic degree number n, compare the degree constituent equation 3.9. Substitution yields

asteosuus-yhtälö

$$k_n = \frac{1}{4\pi} \int_0^{2\pi} \int_{-\pi/2}^{+\pi/2} T_P T_{n,P} \cos\phi \, d\phi \, d\lambda =$$
$$= \frac{1}{4\pi} \iint_\sigma T T_n \, d\sigma = M\{T T_n\} = \frac{1}{4\pi} \iint_\sigma T_n^2 d\sigma = M\{T_n^2\},$$

according to the definition of operator M and considering the mutual orthogonality of the functions T_n.

The degree variances are the geographic variances of the degree constituents of the disturbing potential.

Write, following equation 4.11, but using the definitions of equation 3.15:

$$T(\phi, \lambda, r) =$$
$$= \frac{GM_\oplus}{R} \sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+1} \sum_{m=0}^{n} \overline{P}_{nm}(\sin\phi) \left(\delta\overline{C}_{nm} \cos m\lambda + \overline{S}_{nm} \sin m\lambda\right),$$

in which the normal field, coefficients $\overline{C}_n^*$, has been subtracted out:

$$\begin{cases} \delta\overline{C}_{n0} = \overline{C}_{n0} - \overline{C}_n^* & \text{if } n \text{ even,} \\ \delta\overline{C}_{nm} = \overline{C}_{nm} & \text{otherwise.} \end{cases}$$

We see that

$$T_n(\phi, \lambda) = \frac{GM_\oplus}{R} \sum_{m=0}^{n} \overline{P}_{nm}(\sin\phi) \left(\delta\overline{C}_{nm} \cos m\lambda + \overline{S}_{nm} \sin m\lambda\right).$$

We obtain

$$k_n = \frac{1}{4\pi} \iint_\sigma T_n^2 \, d\sigma = \langle T_n \cdot T_n \rangle_{\overline{\sigma}} = \left(\frac{GM_\oplus}{R}\right)^2 \sum_{m=0}^{n} \left(\delta\overline{C}_{nm}^2 + \overline{S}_{nm}^2\right).$$

Here, we have exploited the orthonormality of the fully normalised basis functions $\overline{P}_{nm}(\sin\phi) \cos m\lambda$ and $\overline{P}_{nm}(\sin\phi) \sin m\lambda$ on the surface of unit sphere σ. So

kantafunktio

The degree variances k_n of the disturbing potential can be calculated directly from the spherical-harmonic coefficients.

The literature offers many alternative notations, such as

$$k_n = \sigma_n^2 = \sigma_i^{TT}.$$

10.8 Propagation of covariances between various quantities

The covariance function K of the disturbing potential derived above can also be used to derive the covariance functions of other quantities. This works in principle for quantities that can be expressed as linear functionals of the disturbing potential $T(\cdot, \cdot, R)$ on the surface of the spherical Earth, as explained in section 10.2.

10.8.1 Example: upwards continuation of the potential

Let us write the disturbing potential in space $T(\phi, \lambda, r)$ as a functional of the surface disturbing potential $T(\phi, \lambda, R) = T(\cdot, \cdot, R)$. With the definition of the degree constituents T_n,

$$T(\phi, \lambda, R) \stackrel{\text{def}}{=} \sum_{n=2}^{\infty} T_n(\phi, \lambda),$$

it holds that

$$T(\phi, \lambda, r) = \sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+1} T_n(\phi, \lambda).$$

Symbolically

$$T(\phi, \lambda, r) = L\{T(\cdot, \cdot, R)\}.$$

Here, L is the linear functional

$$L\{f\} = \sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+1} f_n,$$

in which the f_n are defined according to degree constituent equation 3.9, so that on the sea level of a spherical Earth

$$f = \sum_{n=2}^{\infty} f_n .$$

Symbolically

$$L\{f\} = \sum_{n=2}^{\infty} L^n f_n ,$$

in which

$$L^n = \left(\frac{R}{r}\right)^{n+1}$$

is the *spectral representation* of the functional L.

We may write *at a certain point* P, location (ϕ_P, λ_P, r_P) in space:

$$L_P\{f\} = \sum_{n=2}^{\infty} L_P^n f_{n,P} ,$$

in which

$$L_P^n = \left(\frac{R}{r_P}\right)^{n+1} .$$

Concretely, for the disturbing potential $T(\phi_P, \lambda_P, r_P)$ in point P, this means

$$T(\phi_P, \lambda_P, r_P) = L_P\{T(\cdot,\cdot,R)\} = \sum_{n=2}^{\infty} L_P^n T_{n,P} = \sum_{n=2}^{\infty} \left(\frac{R}{r_P}\right)^{n+1} T_n(\phi_P, \lambda_P).$$

The covariance function *in space* of the disturbing potential T is

$$\begin{aligned} K(r_P, r_Q, \psi_{PQ}) &= M'\Big\{T(\phi_P, \lambda_P, r_P)\, T(\phi_{Q(P)}, \lambda_{Q(P)}, r_Q)\Big\} = \\ &= M'\Big\{L_P\{T(\cdot,\cdot,R)\}\, L_{Q(P)}\{T(\cdot,\cdot,R)\}\Big\} = \\ &= M'\Big\{\sum_{n=2}^{\infty} (L_P^n T_{n,P}) \sum_{n'=2}^{\infty} \left(L_Q^{n'} T_{n',Q(P)}\right)\Big\} = \\ &= \sum_{n=2}^{\infty} \sum_{n'=2}^{\infty} L_P^n L_Q^{n'} M'\{T_{n,P} T_{n',Q(P)}\}. \end{aligned}$$

Based on orthogonality,[10] $M'\{T_{n,P}T_{n',Q(P)}\} = 0$ if $n \neq n'$. So [10]

$$K(r_P, r_Q, \psi_{PQ}) = \sum_{n=2}^{\infty} L_P^n L_Q^n M'\{T_{n,P}T_{n,Q(P)}\}. \tag{10.16}$$

Now at sea level, all $L_P^n = L_Q^n = 1$, so

$$K(\psi_{PQ}) = \sum_{n=2}^{\infty} M'\{T_{n,P}T_{n,Q(P)}\}.$$

Comparing with equation 10.14,

$$K(\psi_{PQ}) = \sum_{n=2}^{\infty} k_n P_n(\cos\psi), \tag{10.14}$$

we see that

$$M'\{T_{n,P}T_{n,Q(P)}\} = k_n P_n(\cos\psi_{PQ}).$$

---

[10] Like in the proof of the harmonicity of $r\Delta g$ in section 8.5, one must take along the third dimension. Write

$$M'\{T_{n,P}T_{n',Q(P)}\} = \frac{1}{2\pi}\int_0^{2\pi} M\{T_{n,P}\, T_{n',Q(P)}\}\, d\alpha_{PQ} =$$
$$= M\left\{T_{n,P}\cdot\frac{1}{2\pi}\int_0^{2\pi} T_{n',Q(P)}\, d\alpha_{PQ}\right\} = M\left\{T_{n,P}\, T^{\bigcirc}_{n',P}\right\},$$

with the definition

$$T^{\bigcirc}_{n',P} \stackrel{\text{def}}{=} \frac{1}{2\pi}\int_0^{2\pi} T_{n',Q(P)}\, d\alpha_{PQ}.$$

With the radial dimension r:

$$T^{\bigcirc}_{n',P}(r) = \frac{1}{2\pi}\int_0^{2\pi} T_{n',Q(P)}(r)\, d\alpha_{PQ} = \frac{1}{2\pi}\int_0^{2\pi}\left(\frac{r}{R}\right)^{n'+1} T_{n',Q(P)}(R)\, d\alpha_{PQ} =$$
$$= \left(\frac{r}{R}\right)^{n'+1}\cdot\frac{1}{2\pi}\int_0^{2\pi} T_{n',Q(P)}(R)\, d\alpha_{PQ} = \left(\frac{r}{R}\right)^{n'+1} T^{\bigcirc}_{n',P}(R).$$

This shows that $T^{\bigcirc}_{n',P}(r)$ is a perfectly legal solid spherical harmonic of degree n', $T^{\bigcirc}_{n',P} = T^{\bigcirc}_{n',P}(R)$ a legal surface harmonic, and the orthogonality of spherical harmonics applies:

$$M'\{T_{n,P}T_{n',Q(P)}\} = M\left\{T_{n,P}T^{\bigcirc}_{n',P}\right\} = 0 \qquad \text{if } n \neq n'.$$

This should not surprise us: if the spatial covariance function is isotropic, it must have the general form

$$K(r_P, r_Q, \psi_{PQ}) = \sum_{n=2}^{\infty} K_n^r(r_P, r_Q)\, K_n^\psi(\psi_{PQ}),$$

and $K_n^\psi(\psi_{PQ})$ must have the same form as $K(\psi)$ in equation 10.14, and for the same reason:

$$K_n^\psi(\psi_{PQ}) = k_n P_n(\cos\psi_{PQ}).$$

At sea level $K_n^r(R, R) = 1$ and the coefficients k_n are those given by equation 10.15.

[11] Equation 10.16 now becomes[11]

$$\begin{aligned} K(r_P, r_Q, \psi_{PQ}) &= \sum_{n=2}^{\infty} L_P^n L_Q^n k_n P_n(\cos\psi_{PQ}) = \\ &= \sum_{n=2}^{\infty} \left(\frac{R}{r_P}\right)^{n+1} \left(\frac{R}{r_Q}\right)^{n+1} k_n P_n(\cos\psi_{PQ}) = \\ &= \sum_{n=2}^{\infty} \left(\frac{R^2}{r_P r_Q}\right)^{n+1} k_n P_n(\cos\psi_{PQ}). \quad (10.17) \end{aligned}$$

Here we have expressed the covariance function of the disturbing potential in space $T(\phi, \lambda, r)$ into an expansion into the degree variances k_n of the corresponding sea-level disturbing potential $T(\phi, \lambda, R)$, by applying *propagation of covariances* on expansion 10.14 of function K. Thus we have obtained the *three-dimensional covariance function* for the disturbing potential, needed for example in mountainous countries and in air and space applications.

kovarianssien kasautuminen

[11] This only works this cleanly because in this case, the operator L^n is of multiplier type, $(R/r)^{n+1}$.

10.8.2 *Example:* the covariance function of gravity anomalies

We know from equation 5.7 that there exists the following relationship between gravity anomalies and the disturbing potential:

$$\Delta g = \sum_{n=2}^{\infty} \frac{n-1}{r} \left(\frac{R}{r}\right)^{n+1} T_n,$$

symbolically: $\Delta g = L_{\Delta g}\{T\}$ for a suitable operator $L_{\Delta g}$:

$$L_{\Delta g}\{f\} = \sum_{n=2}^{\infty} L^n_{\Delta g} f_n, \qquad L^n_{\Delta g} = \frac{n-1}{r} \left(\frac{R}{r}\right)^{n+1}.$$

Again, at a concrete point P,

$$\Delta g(\phi_P, \lambda_P, r_P) = L_{\Delta g,P}\{T(\cdot,\cdot,R)\} = \\ = \sum_{n=2}^{\infty} L^n_{\Delta g,P} T_{n,P} = \sum_{n=2}^{\infty} \frac{n-1}{r_P} \left(\frac{R}{r_P}\right)^{n+1} T_{n,P}.$$

We can show in the same way as above that

$$\mathrm{Cov}\{\Delta g_P, \Delta g_Q\} = M'\{\Delta g_P \, \Delta g_{Q(P)}\} = \\ = \sum_{n=2}^{\infty} L^n_{\Delta g,P} L^n_{\Delta g,Q} M'\{T_{n,P} T_{n,Q(P)}\} = \\ = \sum_{n=2}^{\infty} \frac{n-1}{r_P} \left(\frac{R}{r_P}\right)^{n+1} \frac{n-1}{r_Q} \left(\frac{R}{r_Q}\right)^{n+1} k_n P_n(\cos\psi_{PQ}) = \\ = \sum_{n=2}^{\infty} \left(\frac{R^2}{r_P r_Q}\right)^{n+2} \left(\frac{n-1}{R}\right)^2 k_n P_n(\cos\psi_{PQ}).$$

Often, we write

$$C(r_P, r_Q, \psi_{PQ}) \overset{\text{def}}{=} \mathrm{Cov}\{\Delta g_P, \Delta g_Q\} = M'\{\Delta g_P \, \Delta g_{Q(P)}\} = \\ = \sum_{n=2}^{\infty} \left(\frac{R^2}{r_P r_Q}\right)^{n+2} c_n P_n(\cos\psi_{PQ}),$$

in which the *degree variances of gravity anomalies* are

$$c_n = \left(\frac{n-1}{R}\right)^2 k_n.$$

Similarly we also calculate the "mixed covariances" between disturbing potential and gravity anomaly:

$$\begin{aligned}
&\mathrm{Cov}\{T_P, \Delta g_Q\} = \\
&\quad = M'\{T_P\, \Delta g_{Q(P)}\} = \sum_{n=2}^{\infty} L_P^n L_{\Delta g,Q}^n M'\{T_{n,P} T_{n,Q(P)}\} = \\
&\quad = \sum_{n=2}^{\infty} \left(\frac{R}{r_P}\right)^{n+1} \frac{n-1}{r_Q} \left(\frac{R}{r_Q}\right)^{n+1} k_n P_n(\cos\psi_{PQ}) = \\
&\quad = \sum_{n=2}^{\infty} \frac{n-1}{r_Q} \left(\frac{R^2}{r_P r_Q}\right)^{n+1} k_n P_n(\cos\psi_{PQ}).
\end{aligned}$$

All these are examples of *propagation of covariances*, when applied to a series expansion:

$$\begin{aligned}
\mathrm{Cov}\left\{L_1\{\underline{T}_P\}, L_2\{\underline{T}_Q\}\right\} &= \sum_n L_{1,P}^n L_{2,Q}^n M'\{T_{n,P} T_{n,Q(P)}\} = \\
&= \sum_n L_{1,P}^n L_{2,Q}^n k_n P_n(\cos\psi_{PQ}),
\end{aligned}$$

for *arbitrary* linear functionals

$$L_1\{T_P\} = \sum_n L_{1,P}^n T_{n,P}, \qquad L_2\{T_Q\} = \sum_n L_{2,Q}^n T_{n,Q},$$

in which $T_{n,P} = T_n(\phi_P, \lambda_P)$ and $T_{n,Q} = T_n(\phi_Q, \lambda_Q)$ are the degree constituents of the disturbing potential on the Earth's surface. The problem, in each case, is identifying the spectral form of this linear functional. This is done by expanding the quantity concerned into T_n, and lifting the coefficients found from the expansion. These coefficients are indicated above by red and blue colouring.

10.9 Global covariance functions

Empirical covariance functions have been calculated often. There have been only a few empirical covariance functions for the whole Earth. They are commonly given in the form of a *degree variance formula*. The best known is the rule observed by William Kaula (Rapp, 1989):[12]

$$k_n = \alpha \frac{2n+1}{n^4}.$$

By writing

$$c_n = \left(\frac{n-1}{R}\right)^2 k_n,$$

in which c_n are the degree variances of *gravity anomalies*, we obtain

$$c_n = \alpha \frac{2n+1}{n^4} \left(\frac{n-1}{R}\right)^2 \approx \frac{2\alpha}{nR^2}.$$

Here, α is a planet specific constant, according to Kaula's estimate $\alpha = 10^{-10} \left(GM_\oplus / a_\oplus\right)^2$.

The Kaula rule does not hold very precisely. It applies, by the way, fairly well for the gravity field of Mars, of course with a different constant (Yuan et al., 2001).

Another well-known rule is the Tscherning–Rapp equation (Tscherning and Rapp, 1974):

$$c_n = \frac{A\,(n-1)}{(n-2)\,(n+B)} = \left(\frac{n-1}{R}\right)^2 k_n.$$

The constants are, according to the authors, $A = 425.28\,\mathrm{mGal}^2$ and $B = 24$ (exactly). As a technical detail, one usually chooses $R = R_B = 0.999\overline{R}$, the radius of a Bjerhammar[13] *sphere* inside the Earth ($\overline{R}$ is the Earth's mean radius). The form of the equation is chosen so the covariance functions of various quantities will be closed expressions.

[12] William M. Kaula (1926–2000) was an American geophysicist and space geodesist who studied the determination of the Earth's gravity field by means of satellite geodesy.

[13] Arne Bjerhammar (1917–2011) was a Swedish geodesist.

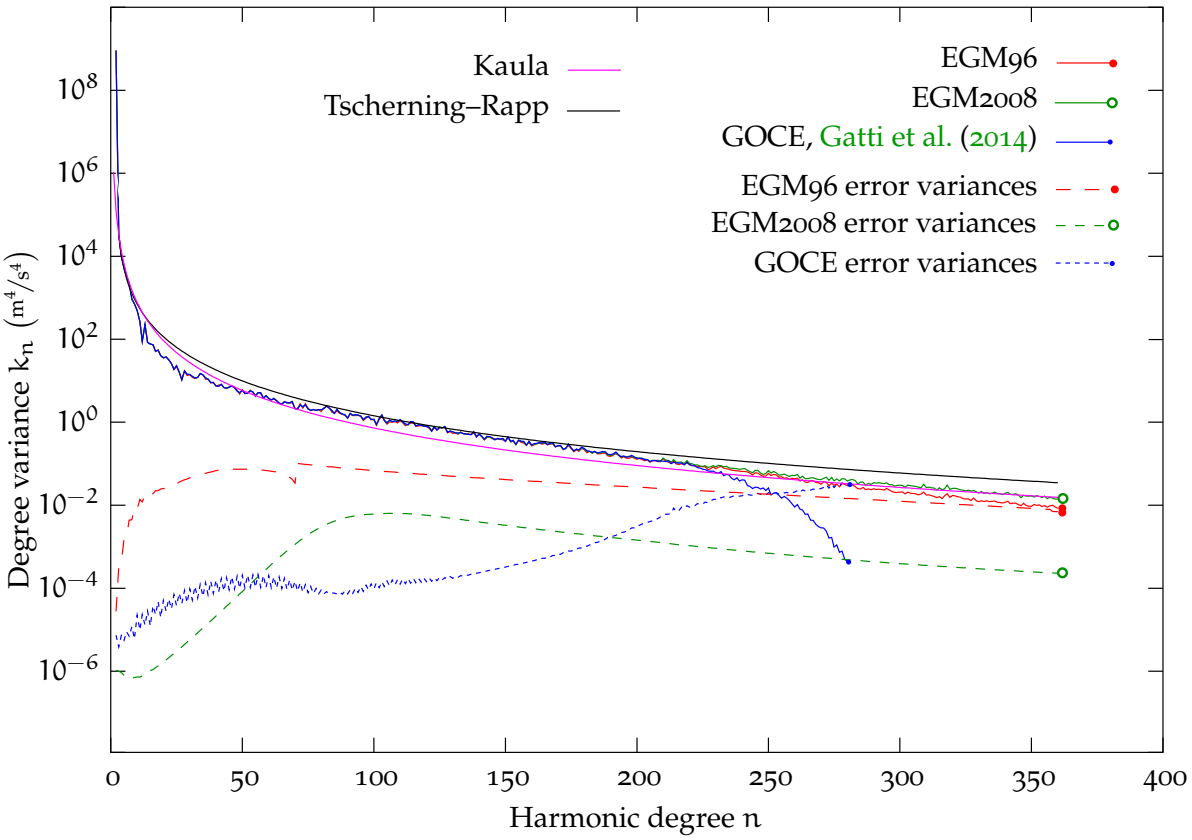


FIGURE 10.5. Global covariance functions as degree variances. The GOCE model cuts off at degree 280.

10.10 Collocation and the spectral viewpoint

The calculations in least-squares collocation can also be executed efficiently by way of FFT. For this one should study the symmetries present in the geometry, especially the *rotational symmetry*, which exists, for example, in the direction of longitude on the whole Earth when collocation equations depend only on *differences* in longitude $\Delta\lambda$ between points, not on absolute longitudes λ.

In the following we discuss a simplified example in one dimension. Let observations $\underline{\ell}_i = \underline{g}_i + \underline{n}_i$ of a field $\underline{g}(\psi)$, $\psi \in [0, 2\pi)$ be given on the edge of a *circle*, in points $\psi_i \overset{\text{def}}{=} 2\pi\, i/N$, $i = 0, 1, 2, \ldots, N-1$. Assume

that also the results of the calculation, estimates $\widehat{f}_i$ of the result function $f(\psi)$, are desired at the same points. Then we have equation 10.9:

$$\widehat{\mathbf{f}} = C_{fg}\left(C_{gg} + D_{gg}\right)^{-1}\left(\underline{\mathbf{g}} + \underline{\mathbf{n}}\right), \tag{10.9}$$

with

$$\begin{aligned}
\left[C_{fg}\right]_{ij} &= C_{fg}\big(f(\psi_i), g(\psi_j)\big) = C_{fg}\big(\psi_i, \psi_j\big),\\
\left[C_{gg}\right]_{ij} &= C_{gg}\big(g(\psi_i), g(\psi_j)\big) = C_{gg}\big(\psi_i, \psi_j\big),\\
\left[D_{gg}\right]_{ij} &= D_{gg}\big(g(\psi_i), g(\psi_j)\big) = D_{gg}\big(\psi_i, \psi_j\big).
\end{aligned}$$

If the *physics* of the whole situation, including the physics of the measurement process, is rotationally symmetric, we must have

$$\left[C_{fg}\right]_{i,j(i)} = M_{\bigcirc}\Big\{f(\psi_i)\, g\big(\psi_{j(i)}\big)\Big\} = \frac{1}{N}\sum_{i=0}^{N-1} f(\psi_i)\, g\big(\psi_{j(i)}\big) \stackrel{\text{def}}{=} \left[C_{fg}\right]_k,$$

with $j(i) = (i + k) \bmod N$. Here, the operator $M_{\bigcirc}$ is the "circle mean" of a function,

$$M_{\bigcirc}\{h\} \stackrel{\text{def}}{=} \frac{1}{N}\sum_{i=0}^{N-1} h(\psi_i),$$

which, like the geographic average in section 10.4, replaces the statistical average.

In the same way we obtain

$$\begin{aligned}
\left[C_{gg}\right]_{i,j(i)} = M_{\bigcirc}\Big\{g(\psi_i)\, g\big(\psi_{j(i)}\big)\Big\} &= \\
&= \frac{1}{N}\sum_{i=0}^{N-1} g(\psi_i)\, g\big(\psi_{j(i)}\big) \stackrel{\text{def}}{=} \left[C_{gg}\right]_k.
\end{aligned}$$

Now C_{fg}, C_{gg} are functions of only k, and we may write them

$$\begin{aligned}
\left[C_{fg}\right]_{ij} &= C_{fg}\big(\psi_i, \psi_j\big) = C_{fg}(\Delta\psi_k) = \left[C_{fg}\right]_k,\\
\left[C_{gg}\right]_{ij} &= C_{gg}\big(\psi_i, \psi_j\big) = C_{gg}(\Delta\psi_k) = \left[C_{gg}\right]_k,
\end{aligned}$$

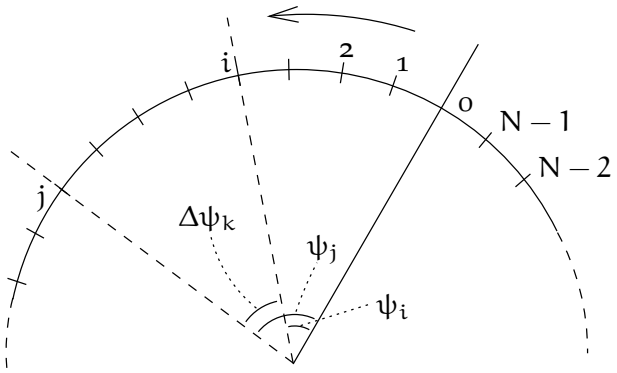


FIGURE 10.6. Circular geometry.

in which $\Delta\psi_k \stackrel{\text{def}}{=} (\psi_j - \psi_i) \bmod 2\pi$ and $k = (j - i) \bmod N$.

Furthermore

$$\left[D_{gg}\right]_{ij} = D_{gg}(\psi_i, \psi_j) = D_{gg}(\Delta\psi_k) = \left[D_{gg}\right]_k = E\{\underline{n}_i \underline{n}_{j(i)}\},$$

the traditional statistical variance of the observation noise. Also because generally the observations do not correlate with each other, we have[14]

$$D_{gg} = \sigma^2 I_N,$$

σ^2 (the variance of observations, assumed equal for all) times the $N \times N$ sized unit matrix.

Toeplitz-sirkulantti matriisi [15]

Matrices of this form are called *Toeplitz circulant* matrices.[15] Thanks to this property, equation 10.9 is a string of convolutions.

[14] In fact, the unit or identity matrix is also known as the Kronecker delta, and as a Toeplitz matrix may be interpreted as a discrete version of the Dirac delta function. Its discrete Fourier transform is "white":

$$\mathcal{F}\{I_N\} = \frac{1}{N},$$

with the same power for all frequencies.

[15] Otto Toeplitz (1881–1940) was a German Jewish mathematician who contributed to functional analysis.

Without proof we present that the spectral version of equation 10.9 looks like this:

$$\mathcal{F}\{\widehat{\mathbf{f}}\} = \frac{\mathcal{F}\{C_{fg}\}}{\mathcal{F}\{C_{gg}\} + \mathcal{F}\{D_{gg}\}}\mathcal{F}\{\underline{\mathbf{g}} + \underline{\mathbf{n}}\} = \frac{\mathcal{F}\{C_{fg}\}}{\mathcal{F}\{C_{gg}\} + \sigma^2/N}\mathcal{F}\{\underline{\mathbf{g}} + \underline{\mathbf{n}}\}.$$

This is an easy and fast way to calculate the solution using FFT. If for a suitable operator L we have $\mathbf{f} = L\{\mathbf{g}\}$, the equation becomes

$$\mathcal{F}\{\widehat{\mathbf{f}}\} = \frac{\mathcal{F}\{L\} \cdot \mathcal{F}\{C_{gg}\}}{\mathcal{F}\{C_{gg}\} + \sigma^2/N} \cdot \mathcal{F}\{\underline{\mathbf{g}} + \underline{\mathbf{n}}\}.$$

In the limit in which the observations are exact, $\sigma^2 = 0$ and thus $\underline{\mathbf{n}} = 0$, it holds that

$$\mathcal{F}\{\widehat{\mathbf{f}}\} = \mathcal{F}\{L\} \cdot \mathcal{F}\{\underline{\mathbf{g}}\} \quad \Longleftrightarrow \quad \widehat{\mathbf{f}} = L\{\underline{\mathbf{g}}\}.$$

For example, if $\mathbf{g}$ are gravity anomalies and $\mathbf{f}$ are values of the disturbing potential, then[16]

$$\mathcal{F}\{L\} = \frac{R}{n-1}.$$

The approach is called *Fast Collocation*, for example Bottoni and Barzaghi (1993). Of course it is used in two dimensions on the Earth's surface, although our example is one-dimensional. As always, it requires that the observations are given on a *grid*, and in this case also that the precision of the material is *homogeneous* — the same everywhere — over the area. This requirement is hardly ever precisely fulfilled.

Self-test questions

1. What is the difference between signal and noise?
2. What is a functional?

[16] In real computation though, the equation must be changed to use, instead of the harmonic degree number n, which refers to global spherical geometry, the Fourier wave number expressed on the computation grid used.

3. What is a linear functional?
4. The statistical behaviour of a stochastic process defined on the Earth's surface is the same independently of where on Earth you are. This property is called isotropy | ergodicity | homogeneity | stationarity.
5. The statistical behaviour of a stochastic process of time is the same independently of where on the time axis you are. This property is called isotropy | ergodicity | homogeneity | stationarity.
6. Why, in the study of the Earth's gravity field, does one use as the average of quantities the geographical average rather than the statistical average?
7. Which two different kinds of covariance functions are used for gravity anomalies on the Earth's surface? Give the formulas and name the free parameters.
8. Explain degree variances. What is the difference between degree variances k_n and c_n?
9. What does Kaula's rule express?
10. What is a Toeplitz circulant matrix?

Exercise 10−1: Variance of prediction

The equation for the variance of prediction at a point P is

$$\Sigma_{PP} = C_{PP} - C_{Pi}(C_{ij} + D_{ij})^{-1}C_{jP},$$

in which the observation points are $i = 1, \ldots, N$. Assume there is only one observation point, point P. Then

$$\Sigma_{PP} = C_{PP} - C_{PP}(C_{PP} + D_{PP})^{-1}C_{PP}.$$

Show that, if $D_{ij} \neq 0$ but however $D_{ij} \ll C_{ij}$,

$$\Sigma_{PP} \approx D_{PP}.$$

Exercise 10–2: Hirvonen's covariance equation and prediction

Hirvonen's covariance equation is

$$C(s) = \frac{C_0}{1 + (s/d)^2}, \tag{10.18}$$

with the Ohio parameters $C_0 = 337\,\text{mGal}^2$ and $d = 40\,\text{km}$ (Heiskanen and Moritz, 1967, equation 7-9). The equation gives the covariance between the gravity anomalies at two points P and Q:

$$C(s_{PQ}) = \text{Cov}\left\{\underline{\Delta g}_P, \underline{\Delta g}_Q\right\}.$$

s_{PQ} is the linear distance between the points.

1. Calculate $\text{Var}\left\{\underline{\Delta g}_P\right\}$ and $\text{Var}\left\{\underline{\Delta g}_Q\right\}$. Remember that, according to the definition, $\text{Var}\{\underline{x}\} = \text{Cov}\{\underline{x}, \underline{x}\}$!
2. Calculate $\text{Cov}\left\{\underline{\Delta g}_P, \underline{\Delta g}_Q\right\}$ if $s_{PQ} = 20\,\text{km}$.
3. Calculate the *correlation*

$$\text{Corr}\left\{\underline{\Delta g}_P, \underline{\Delta g}_Q\right\} \stackrel{\text{def}}{=} \frac{\text{Cov}\left\{\underline{\Delta g}_P, \underline{\Delta g}_Q\right\}}{\sqrt{\text{Var}\left\{\underline{\Delta g}_P\right\}\,\text{Var}\left\{\underline{\Delta g}_Q\right\}}}.$$

4. Assume now that *we only have a measurement in point* P. What is the "variance of prediction" of the gravity anomaly in point Q which is at a distance $s_{PQ} = 10\,\text{km}$ from the (precisely!) given anomaly in point P? Apply equation 10.11 as follows:

$$\Sigma_{QQ} = C_{QQ} - C_{QP} C_{PP}^{-1} C_{PQ}.$$

5. And item 4 if the distance is $s_{PQ} = 80\,\text{km}$?

Exercise 10–3: Predicting gravity anomalies

Let the measured gravity anomalies $\underline{\ell}_1 = \underline{\Delta g}_1 + \underline{n}_1$ and $\underline{\ell}_2 = \underline{\Delta g}_2 + \underline{n}_2$ be given at two points 1 and 2. The distance between the points is 80 km

and between them, at the same distance of 40 km from both, is located point P. Compute the gravity anomaly of point P, Δg_P by means of prediction. The prediction equation is

$$\widehat{\Delta g}_P = C_{Pi} \left(C_{ij} + D_{ij}\right)^{-1} \underline{\ell}_j,$$

where $\underline{\ell}_j = \underline{\Delta g}_j + \underline{n}_j$ is the (abstract) vector of gravity anomaly observations,

$$C_{ij} = \begin{bmatrix} \mathrm{Var}\left\{\underline{\Delta g}_1\right\} & \mathrm{Cov}\left\{\underline{\Delta g}_1, \underline{\Delta g}_2\right\} \\ \mathrm{Cov}\left\{\underline{\Delta g}_1, \underline{\Delta g}_2\right\} & \mathrm{Var}\left\{\underline{\Delta g}_2\right\} \end{bmatrix}$$

is the signal variance matrix of the vector $\underline{\Delta g}_i$, and

$$C_{Pi} = \begin{bmatrix} \mathrm{Cov}\left\{\underline{\Delta g}_P, \underline{\Delta g}_1\right\} & \mathrm{Cov}\left\{\underline{\Delta g}_P, \underline{\Delta g}_2\right\} \end{bmatrix}$$

is the signal covariance matrix between $\underline{\Delta g}_P$ and $\underline{\Delta g}_i$. D_{ij} is the variance matrix of the observation random uncertainty or *noise* $\underline{n}_i$, $i = 1, 2$:

$$D_{ij} = \begin{bmatrix} \mathrm{Var}\{\underline{n}_1\} & \mathrm{Cov}\{\underline{n}_1, \underline{n}_2\} \\ \mathrm{Cov}\{\underline{n}_1, \underline{n}_2\} & \mathrm{Var}\{\underline{n}_2\} \end{bmatrix}.$$

1. Compute the matrix C_{ij}, assuming again Hirvonen's covariance equation 10.18 and a parameter value of $d = 40$ km.
2. Compute C_{Pi}.
3. Compute $\widehat{\Delta g}_P$ expressed in the observed values $\underline{\ell}_1$ and $\underline{\ell}_2$. Assume $D_{ij} = 0$ (and thus $\underline{n}_i = 0$). Inverting the C_{ij} matrix is possible by hand, but just use Matlab or octave.
4. Compute the *variance of prediction* (note $C_{jP} = C_{Pi}^T$) using

$$\Sigma_{PP} = C_{PP} - C_{Pi} C_{ij}^{-1} C_{jP}.$$

Exercise 10−4: Predicting gravity anomalies (2)

Let us again have points 1 and 2 with measured gravity anomalies $\underline{\ell}_1 = \underline{\Delta g}_1$ and $\underline{\ell}_2 = \underline{\Delta g}_2$. Now however the points 1, 2 and P are in a

triangular configuration, with a right angle at point P, and the distances from P to points 1 and 2 still 40 km. The distance between points 1 and 2 is now only $40\sqrt{2}$ km.

1. Compute C_{ij}, C_{Pi}, $\widehat{\Delta g}_P$ and Σ_{PP}.
2. Compare the result with the previous one. Conclusion?

Exercise 10–5: Propagation of covariances

Given the covariance function 10.17 of the disturbing potential

$$\mathrm{Cov}\{\underline{T}_P, \underline{T}_Q\} = \sum_{n=2}^{\infty} \left(\frac{R^2}{r_P r_Q}\right)^{n+1} k_n P_n(\cos\psi_{PQ}).$$

1. Derive the covariance function of the gravity disturbance $\underline{\delta g}$ (equation 5.4). Hint: write first an expansion of the form

$$\delta g = \sum_{n=2}^{\infty} L^n_{\delta g} T_n$$

in order to find the expression for the coefficient $L^n_{\delta g}$. After this

$$\mathrm{Cov}\left\{\underline{\delta g}_P, \underline{\delta g}_Q\right\} = \sum_{n=2}^{\infty} L^n_{\delta g,P} L^n_{\delta g,Q} k_n P_n(\cos\psi_{PQ}).$$

2. Derive the covariance function of the *gravity-gradient disturbance*

$$\frac{\partial^2}{\partial r^2} T = -\frac{\partial}{\partial r}\delta g,$$

in other words, the vertical gradient of the gravity *disturbance*.

Exercise 10–6: Kaula's rule for gravity gradients

For the disturbing potential

$$T(\phi, \lambda, r) = \sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+1} T_n(\phi, \lambda)$$

or on the Earth's surface ($r = R$):

$$T(\phi, \lambda, R) = \sum_{n=2}^{\infty} T_n(\phi, \lambda)$$

Kaula's rule applies, with the *degree variances*

$$k_n = \alpha \frac{2n+1}{n^4}.$$

From these one can derive, using propagation of variances, the degree variances of *gravity anomalies*

$$\Delta g(\phi, \lambda, R) = \sum_{n=2}^{\infty} L^n_{\Delta g}(R)\, T_n(\phi, \lambda) = \sum_{n=2}^{\infty} \left(\frac{n-1}{R}\right) T_n(\phi, \lambda)$$

as follows:

$$c_n = \left(L^n_{\Delta g}(R)\right)^2 k_n = \left(\frac{n-1}{R}\right)^2 k_n \approx \frac{2\alpha}{nR^2}.$$

Analogously, differentiate expansion 5.7 for the gravity anomaly

$$\Delta g(r, \varphi, \lambda) = \sum_{n=2}^{\infty} \frac{n-1}{r} \left(\frac{R}{r}\right)^{n+1} T_n(\phi, \lambda),$$

yielding

$$\frac{\partial\, \Delta g}{\partial r} = -\sum_{n=2}^{\infty} \frac{(n-1)\,(n+2)}{r^2} \left(\frac{R}{r}\right)^{n+1} T_n,$$

the connection between the disturbing potential and the *anomalous gravity gradient* in the spectral domain.

On the Earth's surface $r = R$:

$$\left.\frac{\partial\, \Delta g}{\partial r}\right|_{r=R} = -\sum_{n=2}^{\infty} \frac{(n-1)\,(n+2)}{R^2} T_n \overset{\text{def}}{=} \sum_{n=2}^{\infty} L^n_{\Delta g'}(R)\, T_n(\phi, \lambda)$$

with

$$L^n_{\Delta g'}(R) = -\frac{(n-1)\,(n+2)}{R^2}.$$

1. Derive an approximate equation for the degree variances for the *anomalous gravity gradient*. Designate them with the symbol c'_n, in an analogue fashion as above for the gravity anomaly degree variances c_n:

   $$c'_n = x(n) \cdot k_n \approx y \cdot n^z.$$

   Find the expression $x(n)$ and the constants y and z for the case of the Earth.

2. Conclusion?

Exercise 10–7: Underground mass points

1. If a mass point is placed inside the Earth at a depth D beneath an observation point P, what then is the *correlation length* s of the change in ambient gravity g it causes on the Earth's surface, for which $C(s) = \frac{1}{2}C_0$?
2. Thus, if we wish to build a model made of mass points, in which under each observation point Δg_P there is one mass point, how deep should we place them if the correlation length d is given?

Gravimetric measurement devices

11

11.1 History

The first measurement device ever built based on a *pendulum* was a clock. The pendulum equation,

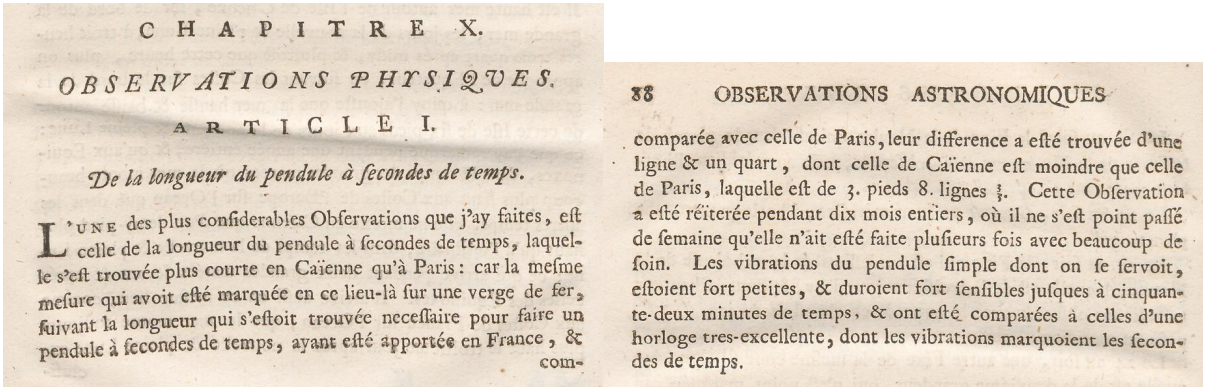
$$P = 2\pi\sqrt{\frac{\ell}{g}},$$

tells that the swinging time or period P of a pendulum of a given length is a constant that depends only on the length ℓ and local gravity g, on condition that the swings are small. The Dutch Christiaan Huygens[1] built in 1657 the first useable pendulum clock based on this principle (Wikipedia, Pendulum clock).

When the young French researcher Jean Richer[2] visited French Guyana in 1671 with a pendulum clock, he noticed that the clock ran clearly slower. The matter was corrected simply by shortening the pendulum. The cause of the effect could not be the climatic conditions in the tropics, like the thermal expansion of the pendulum. The right

[1] Christiaan Huygens FRS (1629–1695) was a leading Dutch natural scientist and mathematician. Besides inventing the pendulum clock, he also was the first to realise (in 1655) that the planet Saturn has a ring.

[2] Jean Richer (1630?–1696) was a French astronomer. He is really only remembered for his pendulum finding.



CHAPITRE X.

OBSERVATIONS PHYSIQUES.

ARTICLE I.

De la longueur du pendule à ſecondes de temps.

L'UNE des plus conſiderables Obſervations que j'ay faites, eſt celle de la longueur du pendule à ſecondes de temps, laquelle s'eſt trouvée plus courte en Caïenne qu'à Paris: car la meſme meſure qui avoit eſté marquée en ce lieu-là ſur une verge de fer, ſuivant la longueur qui s'eſtoit trouvée neceſſaire pour faire un pendule à ſecondes de temps, ayant eſté apportée en France, & com-

88 OBSERVATIONS ASTRONOMIQUES

comparée avec celle de Paris, leur difference a eſté trouvée d'une ligne & un quart, dont celle de Caïenne eſt moindre que celle de Paris, laquelle eſt de 3. pieds 8. lignes ¼. Cette Obſervation a eſté réïterée pendant dix mois entiers, où il ne s'eſt point paſſé de ſemaine qu'elle n'ait eſté faite pluſieurs fois avec beaucoup de ſoin. Les vibrations du pendule ſimple dont on ſe ſervoit, eſtoient fort petites, & duroient fort ſenſibles juſques à cinquante-deux minutes de temps, & ont eſté comparées à celles d'une horloge tres-excellente, dont les vibrations marquoient les ſecondes de temps.

FIGURE 11.1. Jean Richer's report.

explanation was that in the tropics, gravity g is weaker than in Europe. After his return to France in 1673, Richer had to make his pendulum longer again. The observation is described in just one paragraph in his report *Observations astronomiques et physiques faites en l'isle de Caïenne*, Richer (1731), pages 87–88.

This is how the *pendulum gravimeter* was invented. Later, much more precise special devices were built, for example Kater's[3] reversion pendulum, and the four-pendulum Von Sterneck[4] device, which was also used in Finland in the 1920s and 1930s (Pesonen, 1930; Hirvonen, 1937). We must also mention the submarine measurements, including in the Java Sea by the Dutch Vening Meinesz, in which it was observed that above the trenches in the ocean floor there is a notable shortage of gravity, and that they thus are in a state of strong isostatic disequilibrium (Vening Meinesz, 1928).

syvänmeren hauta

Pendulum gravimeters are however too hard to operate and too slow for high-productivity gravimetric observations. For that purpose, the *spring gravimeter* was developed, see section 11.2.

[3] Henry Kater FRS FRAS (1777–1835) was a British physicist who made contributions to scientific instruments and metrology.

[4] Robert Freiherr (baron) Daublebsky von Sterneck (1839–1910) was a major general in the Austro-Hungarian army and a geophysicist, astronomer and geodesist.

FIGURE 11.2. Autograv™ CG-5 spring gravimeter from Scintrex. Image Monniaux (2011).

A pendulum gravimeter is in principle an absolute measurement device, as gravity is obtained directly as an acceleration. There are, however, systematic effects associated with the suspension or *pivot*, of the pendulum, because of which one cannot trust in the absoluteness of the measurement after all. One attempted solution is the *very long wire pendulum*, for example Hytönen (1972). However, nowadays absolute measurements are made with ballistic gravimeters, see section 11.3. It has been observed that the older measurement values obtained with pendulum apparatus in the so-called Potsdam system are systematically 14 mGal too high.

11.2 The relative or spring gravimeter

A spring gravimeter is at its simplest the same as a *spring balance*.

In a linear spring balance the equation of motion of the test mass is

$$m\left(\frac{d^2\ell}{dt^2} - g\right) = -k\left(\ell - \ell_0\right), \tag{11.1}$$

where m is the test mass, g the local (to be measured) gravity, and k the spring constant. The quantity ℓ_0 is the "rest length" of the spring; the length it would have if no external forces were acting on it. ℓ is the true, instantaneous length of the string.

The equilibrium between the spring force and gravity is

$$\frac{d^2\ell}{dt^2} = 0 \quad \Longrightarrow \quad mg = k\left(\ell - \ell_0\right) = k\left(\bar{\ell} - \ell_0\right), \tag{11.2}$$

in which $\bar{\ell}$ is the mean length of the spring during the oscillation, and also the *equilibrium length* in the absence of oscillations.

värähtely-yhtälö

When the test mass is disturbed, it starts oscillating about its equilibrium position. The oscillation equation, obtained by summing equations 11.1 and 11.2, is

$$\frac{d^2}{dt^2}\left(\ell - \bar{\ell}\right) = -\frac{k}{m}\left(\ell - \bar{\ell}\right).$$

The period is

$$P = 2\pi\sqrt{\frac{m}{k}} = 2\pi\sqrt{\frac{\bar{\ell} - \ell_0}{g}} = 2\pi\sqrt{\frac{\delta\ell}{g}}, \tag{11.3}$$

in which $\delta\ell = \bar{\ell} - \ell_0$ denotes the difference between the equilibrium length and the length in the state of rest: the *lengthening of the spring by gravity*.

The *sensitivity* of the instrument is obtained by differentiating equation 11.2 in the form

$$mg = k\left(\bar{\ell} - \ell_0\right) = k\,\delta\ell$$

with the result

$$\frac{d\bar{\ell}}{dg} = \frac{d\left(\delta\ell\right)}{dg} = \frac{m}{k} = \frac{P^2}{4\pi^2}. \tag{11.4}$$

Substitution of, for example, $\delta\ell = 5\,\text{cm}$ and $g = 10\,\text{m/s}^2$ into equation 11.3 yields $P = 0.44\,\text{s}$. One milligal of change in gravity g produces,

according to equation 11.4, a lengthening of only $5 \cdot 10^{-8}\,\mathrm{m} = 50\,\mathrm{nm}$ (check!), one-twelfth the wavelength of a helium-neon laser. Clearly then, the sensor observing or compensating this displacement must be extremely sensitive!

11.2.1 Astatisation

An *astatised gravimeter* uses a different measurement geometry. The LaCoste-Romberg gravimeter, which long enjoyed great popularity, serves as an example. Inside it, the test mass is at the end of a lever beam, see figure 11.3. Two torques operate on the beam, which are in equilibrium. The torque caused by the spring is

$$\tau_s = k\left(\bar{\ell} - \ell_0\right) b \sin\beta,$$

in which $\bar{\ell}$ is the spring's true, stretched equilibrium length, and ℓ_0 the theoretical length without loading, the rest length.

According to the sine rule

$$\bar{\ell}\sin\beta = c\sin(90^\circ + \epsilon) = c\cos\epsilon,$$

substitution of which in the previous equation yields

$$\tau_s = k\left(\bar{\ell} - \ell_0\right)\frac{bc}{\bar{\ell}}\cos\epsilon.$$

The force of gravity pulling at the mass is mg, and the corresponding torque

$$\tau_g = mgp\cos\epsilon.$$

Between these there has to be equilibrium:

$$\tau_g - \tau_s = mgp\cos\epsilon - k\left(\bar{\ell} - \ell_0\right)\frac{bc}{\bar{\ell}}\cos\epsilon = 0$$

or

$$mgp\bar{\ell} - kbc\left(\bar{\ell} - \ell_0\right) = 0. \tag{11.5}$$

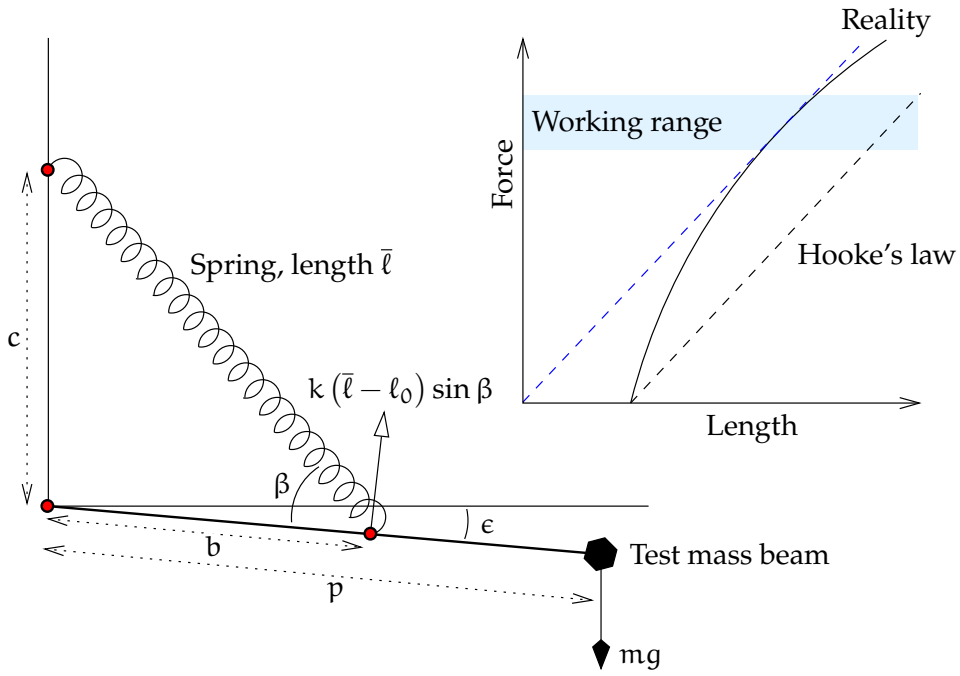


FIGURE 11.3. Operating principle of a spring gravimeter. On the right, how to build a "zero-length spring".

By differentiation

$$mp\bar{\ell}\,dg + mgp\,d\bar{\ell} - kbc\,d\bar{\ell} = 0,$$

from which is obtained, by substituting equation 11.5, a *sensitivity equation*:

$$\frac{d\bar{\ell}}{dg} = -\frac{mp\bar{\ell}}{mgp - kbc} = -\frac{mp\bar{\ell}}{mgp - mgp\,\bar{\ell}\Big/(\bar{\ell} - \ell_0)} = \frac{\bar{\ell}}{g}\frac{\bar{\ell} - \ell_0}{\ell_0}.$$

From this we see that the sensitivity can be driven up arbitrarily by choosing ℓ_0 as short as possible, almost zero — a so-called *zero-length spring* solution (Wikipedia, Zero-length springs).

tasaus rasiatasain Of course, *levelling* the instrument, using its bull's-eye level and three footscrews, is critical.

For example, assuming $\bar{\ell} = 5\,\text{cm}$, $\ell_0 = 0.1\,\text{cm}$, $g = 10\,\text{m/s}^2$ gives

$$\frac{d\bar{\ell}}{dg} = 2.5 \cdot 10^{-6}\,\text{m/mGal},$$

a 50 times[5] better result than earlier! The improvement or *astatisation ratio* is precisely $(\bar{\ell} - \ell_0)/\ell_0$.

This is the operating principle of an *astatised* gravimeter, like the LaCoste-Romberg.[6]

11.2.2 *Period of oscillation*

There is another way to look at this: if the instrument is not in equilibrium, the lever beam will slowly oscillate about the equilibrium position. We start from equation 11.5:

$$mgp\bar{\ell} - kbc\left(\bar{\ell} - \ell_0\right) = 0, \tag{11.6}$$

but for a state of disequilibrium. Then, the test mass will be undergoing an acceleration a, positive downwards, and we have

$$m\,(g - a)\,p\ell - kbc\,(\ell - \ell_0) = 0,$$

where, instead of the equilibrium spring length $\bar{\ell}$, the instantaneous length ℓ appears. Subtracting the above two equations yields

$$mgp\left(\bar{\ell} - \ell\right) + map\ell - kbc\left(\bar{\ell} - \ell\right) = 0.$$

Use equation 11.6 to eliminate kbc, yielding

$$mgp\left(\bar{\ell} - \ell\right) + map\ell - mgp\frac{\bar{\ell}}{\bar{\ell} - \ell_0}\left(\bar{\ell} - \ell\right) = 0.$$

---

[5] For comparability we should still multiply by $p/b\sin\beta$, if the position of the test mass is measured.

[6] Lucien LaCoste (1908–1995) was an American physicist and metrologist, who, as an undergraduate, together with his physics professor Arnold Romberg (1882–1974) discovered the principle of the astatised gravimeter and zero-length spring.

Rearranging terms gives

$$ma\ell = mgp\frac{\ell_0}{\bar{\ell}-\ell_0}\left(\bar{\ell}-\ell\right)$$

or

$$a = -\frac{g}{\ell}\frac{\ell_0}{\bar{\ell}-\ell_0}\left(\ell-\bar{\ell}\right).$$

Here we see again the "astatisation ratio" $\left(\bar{\ell}-\ell_0\right)/\ell_0$ appear, which for a zero-length spring ($\ell_0 \approx 0$) is very large.

Now the string length disequilibrium $\ell - \bar{\ell}$ is connected with the vertical displacement z (reckoned downwards) of the test mass, as follows:

$$z = \left(\ell-\bar{\ell}\right)\frac{p}{b\sin\beta}.$$

With this, we obtain

$$a = \frac{d^2}{dt^2}z = -\frac{g}{\ell}\frac{\ell_0}{\bar{\ell}-\ell_0}\frac{b\sin\beta}{p}z.$$

This is an oscillation equation in z. The oscillation period is

$$P = 2\pi\sqrt{\frac{\ell}{g}\frac{p}{b\sin\beta}\frac{\bar{\ell}-\ell_0}{\ell_0}}.$$

For the same values as above, $\ell_0 = 0.1\,\mathrm{cm}$, $\bar{\ell} = 5\,\mathrm{cm} \approx \ell$, $g = 10\,\mathrm{m/s^2}$, and $p/b\sin\beta = 2$, we find

$$P = 4.4\,\mathrm{s}.$$

What this long oscillation period also means is that the instrument is less sensitive to high-frequency vibrations, for example from passing traffic or microseismicity. This is a significant operational advantage.

11.2.3 *Practicalities of measurement*

An ordinary spring gravimeter is based on *elasticity*. Because there is no material that is perfectly elastic, but it is always plastic[7] as well, the

[7]

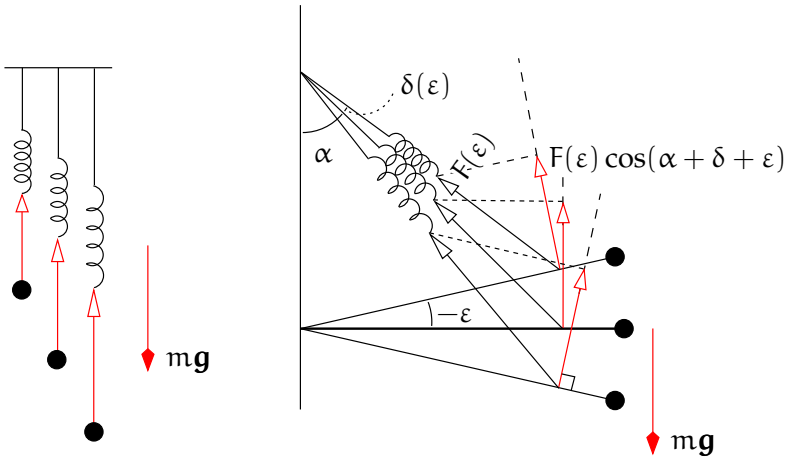


FIGURE 11.4. The idea of astatisation. The elastic force of an ordinary spring grows steeply with extension (left), whereas the weight of the test mass is constant. The lever beam and diagonal arrangement (right) causes the part of the force of the spring in the direction of motion of the lever (red) to diminish with extension, while the spring force itself grows almost similarly with extension. This near-cancellation boosts sensitivity. The spring used is a *zero-length spring*.

gravimeter itself changes during the measurement process. This change is called *drift*. The drift is managed in practical measurements by the following measures: käynti

- We measure along lines starting from a known point and ending on a known point, producing a *closing error*. A line is traversed as rapidly as possible. The closing error is eliminated by adjusting

[7] Plastic deformation in a metal crystal is mediated by crystal-lattice defects called *dislocations*. As dislocations travel through the crystal lattice under load, the properties of the metal change, which may eventually result in *metal fatigue*, a known problem for example in aviation. Wikipedia, Dislocation. The art of making metals stronger by inhibiting the motion of dislocations, for example by adding carbon to iron to form steel, forms a large part of practical metallurgy. Wikipedia, Strengthening mechanisms of materials.

the values obtained from the measurement in proportion to their times of measurement.

- The gravimeter is transported carefully without bumping it.
- We remember to always *arrest* (clamp down the lever beam) during transport!
- Because the elastic properties of the spring and the instrument geometry both depend on temperature, precision gravimeters are always *thermostated*.

arretointi

A sea gravimeter differs from an ordinary (land) gravimeter in having powerful *damping*. This applies also for an airborne gravimeter. Both types are mounted on a stabilised platform, keeping the axis of measurement along the local vertical in spite of vehicle motion.

vaimennus
ilmagravimetri

11.3 The absolute or ballistic gravimeter

The ballistic or absolute gravimeter is a return to roots, the definition of gravity: it measures directly the *acceleration of free fall*. The instrument contains a vacuum tube, inside of which an object, a prism reflecting light, falls freely. See figure 11.5.

Here we describe briefly the JILA gravimeter, built at the University of Colorado at Boulder by Jim Faller's group,[8] of which the Finnish Geodetic Institute has acquired two. Figure 11.6 shows the newer model, FG5, built by the same group. In Finland this instrument, serial number 221, has served as the national standard for the acceleration of free fall. It was upgraded to model FG5X in 2012.

8

During the fall of the prism, a "cage" with a window in the bottom moves along with the prism inside it without touching it. The main

[8] James E. Faller (born 1934) is an American physicist, metrologist, geodesist, and student of gravitation. He proposed the installation of laser retroreflectors on the lunar surface in the context of the Apollo project, in order to measure the distance to the Moon — LLR, lunar laser ranging.

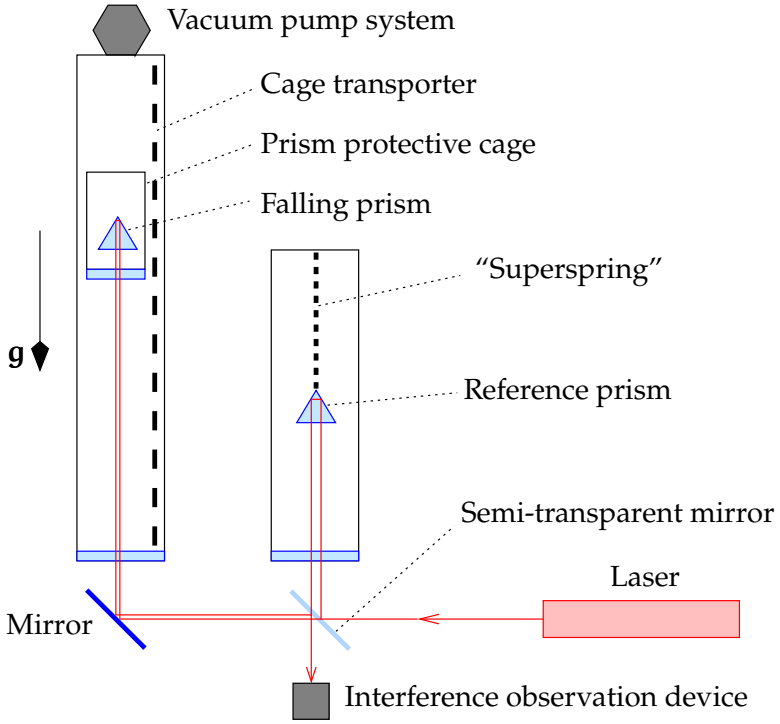


FIGURE 11.5. Operating principle of a ballistic absolute gravimeter.

purpose of the cage is to prevent the last remaining traces of air from affecting the motion of the prism. Approaching the bottom, the cage, which moves along a rail under computer control, decelerates, and the prism lands relatively softly on its base. After that, the cage moves back to the top of the tube and a new measurement cycle starts.

A laser interferometer measures the locations of the prism during its fall. The measurements are repeated thousands of times to get good precision through averaging. Another prism, the reference prism, is suspended in another tube from a very soft spring (actually an electronically simulated "superspring") to protect it from microseismicity.

vertausprisma

The instrument is designed to achieve the greatest precision possible; for example, the vibration caused by the drop is controlled by a well-designed mount. Precisions are of the order of several microgals, similar to what LaCoste-Romberg relative gravimeters are capable of.

FIGURE 11.6. Absolute gravimeter of type FG5. Photograph United States National Oceanic and Atmospheric Administration (NOAA).

The instrument is however large and, although transportable, cannot be called a field instrument. Of late, development has gone in the direction of smaller devices, which are essentially better portable.

The motion of a freely falling mass is given by the equation

$$\frac{d^2}{dt^2}z = g(z),$$

where it is assumed — realistically — that gravity g depends on the location z within the drop tube, reckoned downwards. If we nevertheless take g to be constant, we obtain by integration

$$\frac{d}{dt}z = v_0 + gt, \qquad z = z_0 + v_0 t + \tfrac{1}{2}gt^2,$$

from which we obtain the *observation equations* of the measurement

process

$$\underline{z}_i = \begin{bmatrix} 1 & t_i & \frac{1}{2}t_i^2 \end{bmatrix} \cdot \begin{bmatrix} z_0 \\ v_0 \\ g \end{bmatrix} + \underline{n}_i .$$

Here, the unknowns[9] to be estimated are z_0, v_0 and g. The quantities $\underline{z}_i$ are the interferometrically measured vertical locations of the falling prism, and $\underline{n}_i$ are the measurement errors or "noise". Determining precisely the corresponding measurement time or *epoch* t_i reckoned from a moment close to that of release of the prism is of course essential. The number of measurement values obtained from each individual drop is substantial.

We write the observation equations in matric form:

$$\underline{\boldsymbol{\ell}} = A\mathbf{x} + \underline{\mathbf{n}},$$

in which

$$\underline{\boldsymbol{\ell}} = \begin{bmatrix} \underline{z}_1 \\ \underline{z}_2 \\ \vdots \\ \underline{z}_n \end{bmatrix}, \quad \underline{\mathbf{n}} = \begin{bmatrix} \underline{n}_1 \\ \underline{n}_2 \\ \vdots \\ \underline{n}_n \end{bmatrix}, \quad A = \begin{bmatrix} 1 & t_1 & \frac{1}{2}t_1^2 \\ 1 & t_2 & \frac{1}{2}t_2^2 \\ \vdots & \vdots & \vdots \\ 1 & t_n & \frac{1}{2}t_n^2 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} z_0 \\ v_0 \\ g \end{bmatrix}.$$

The solution follows from this according to the method of least-squares adjustment, from the *normal equations* pienimmän neliösumman tasoitus

$$A^T A\,\hat{\mathbf{x}} = A^T \underline{\boldsymbol{\ell}},$$

giving the solution, or estimate,

$$\hat{\mathbf{x}} = \left(A^T A\right)^{-1} A^T \underline{\boldsymbol{\ell}}.$$

The uncertainty of the estimates is given by the variance matrix

$$\mathrm{Var}\{\hat{\mathbf{x}}\} = \sigma^2 \left(A^T A\right)^{-1},$$

[9] It would be easy (exercise!) to add an unknown representing the vertical gradient of gravity to this. Whether a useable value for this unknown can be obtained from the measurements is a good question.

in which σ is the uncertainty (mean error) of a single observation $\underline{z}_i$, also known as the mean error of unit weight.

painoyksikön keskivirhe

An alternative type of absolute gravimeter *throws* the prism *up* (inside the tube), after which it moves along a symmetric path. An example of such a "rise-and-fall" instrument is the Italian IMGC-02 (d'Agostino et al., 2008). Theoretically, this method would give more precise results; however, the technical challenges are larger than in the case of the dropping method. Intercomparisons between instruments of these two types have helped to identify error sources.

Recently so-called atomic or quantum gravimeters have also been built, in which the falling of individual atoms is measured interferometrically (de Angelis et al., 2009).

The idea of the device is that it measures the effect of gravity on the phase angle of the matter wave of falling atoms. Firstly a so-called Bose–Einstein condensate is prepared by extreme cooling. The condensate consists of perhaps a million atoms in identical quantum states, with their phase angles in lockstep like marching soldiers.

The interaction between laser light and the atoms is based on the Raman effect, an inelastic scattering process, in which the atoms exchange both energy and momentum with photons in the laser beam, while switching from one quantum state to another. The scattering involves two photons: the transition is "forbidden", with a very precisely defined energy and momentum change.

The condensate is dropped, and the first laser pulse splits it into two. Half of the atoms[10] fall first slowly, then faster. The other half fall fast at first and then slower. To bring this about, a second laser pulse pair is used that acts like a mirror, or perhaps a tennis racket. The third and last laser pulse recombines the beams. Then, constructive or destructive interference is observed using a fluorescence detector. The

10

[10] This is a quantum theoretically erroneous statement. The matter wave of *each individual atom* splits into two! Wikipedia, Double-slit experiment.

phase difference between the two arms of the interferometer is inferred from these observations.

As the atoms travel through space-time along two different paths on which the gravity potentials are different,[11] a phase difference is formed between these, which in principle can be measured. See figure 11.7, in which the horizontal axis is time. Without gravity (dashed lines) this phase difference would be zero. 11

Like in all (non-kinematic) interferometric methods, the ambiguity problem — the circumstance that a measured phase is always in the interval $[0, 2\pi)$, although a phase change or difference may contain any number of whole cycles — poses its own challenge. Ambiguity resolution is possible by measuring at several different inter-pulse time intervals T, figure 11.7.

11.4 Network hierarchy in gravimetry

Network hierarchy is just as important in gravimetry as in measurements of location or height. The procedure has typically been that the highest measurement order consisted of points measured by absolute gravimeters — in the old days, this meant pendulum measurements. Stepwise densification of this highest-order network, *base network* measurement, was subsequenty done with relative or spring gravimeters, like also the lowest-order measurements or gravity mapping surveys. In base network measurement, fast transportation was used, such as aircraft, and national or regional reference points were often located at airports. vertauspiste

Because pendulum instruments are not sufficiently precisely absolute, the old Potsdam system collected a 14 mGal systematic error: all values were that much too high. Nowadays we use ballistic free-fall gravimeters instead, the possible systematics of which are much smaller — but

[11] In fact, the spinning of the phase angle of the atom's wave function acts like a clock, and the speed at which time elapses depends on the local geopotential (Vermeer, 1983a).

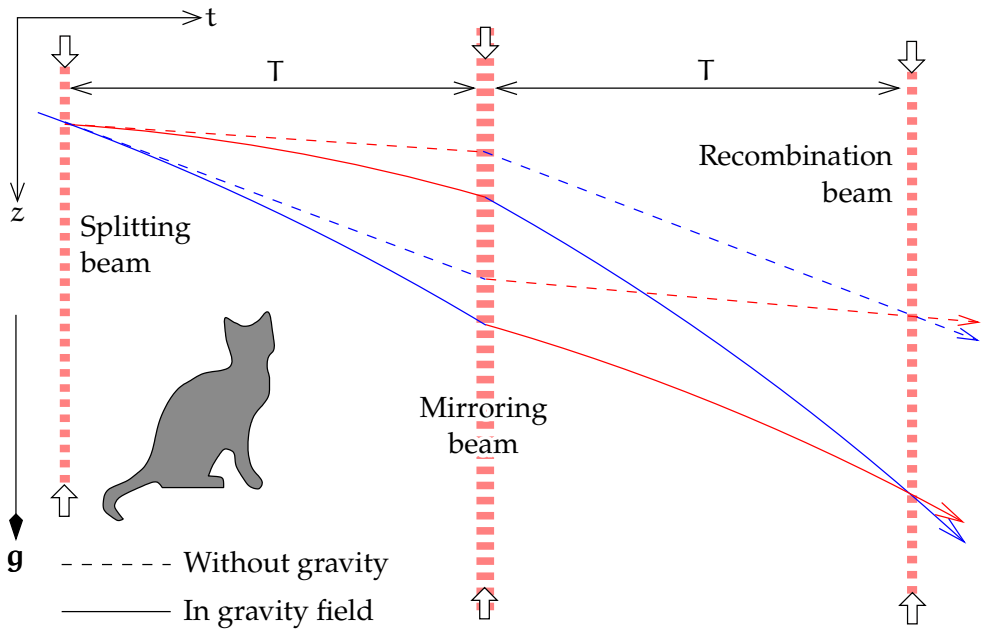


FIGURE 11.7. The idea of operation of an atomic or quantum gravimeter. The horizontal axis is time.

by no means non-existent, in the order of microgals. As there are no better, meaning more absolute, instruments than these, the issue cannot ultimately be resolved. For this reason, international instrument intercomparisons, like the International Intercomparison of Absolute Gravimeters, are regularly organised and are valuable.

In Finland, regular absolute gravimetric measurements have been made in Metsähovi, and also in Vaasa (two points), Joensuu (two points), Kuusamo, Sodankylä, Kevo, and Eurajoki.

11.5 The superconducting gravimeter

This gravimeter type is based on a superconducting metal sphere levitating on a magnetic field. The precise place of the sphere is measured electronically. Because a superconducting material is impenetrable by a magnetic field, the sphere will remain forever in the same spot inside the field:

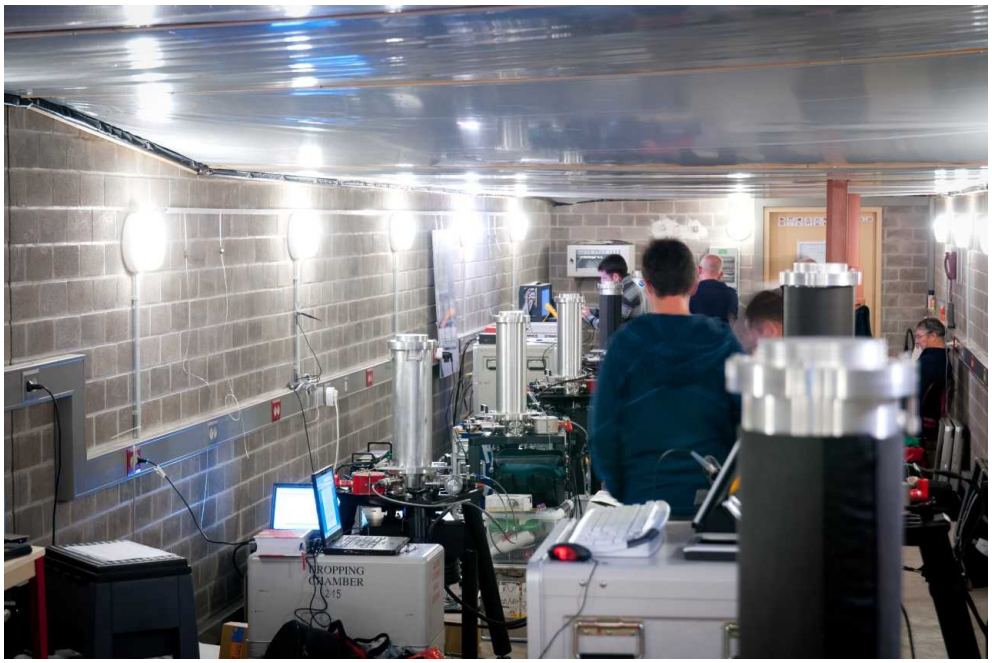


FIGURE 11.8. International intercomparison of absolute gravimeters in Walferdange, Luxembourg. Image courtesy © Olivier Francis.

the Meissner effect. Of course the field itself must be constant. It is generated by superconducting solenoids inside a vessel made of mu-metal käämi (Wikipedia, Mu-metal), which keeps out the Earth's magnetic field.

Superconduction applied for this still demands working at the temperature of liquid helium (He). For this reason the device is not only expensive, but also requires an expensive laboratory in an environment where the societal infrastructure allows.

There are more than thirty superconducting gravimeters in the world. The work is co-ordinated by the IAG service IGETS, the International Geodynamics and Earth Tide Service. One GWR20-type instrument has been working since 1994 in Kirkkonummi at the Metsähovi research station of the then Finnish Geodetic Institute, now the National Land Survey, Virtanen and Kääriäinen (1995), Virtanen (1998). The instrument was upgraded in 2014.

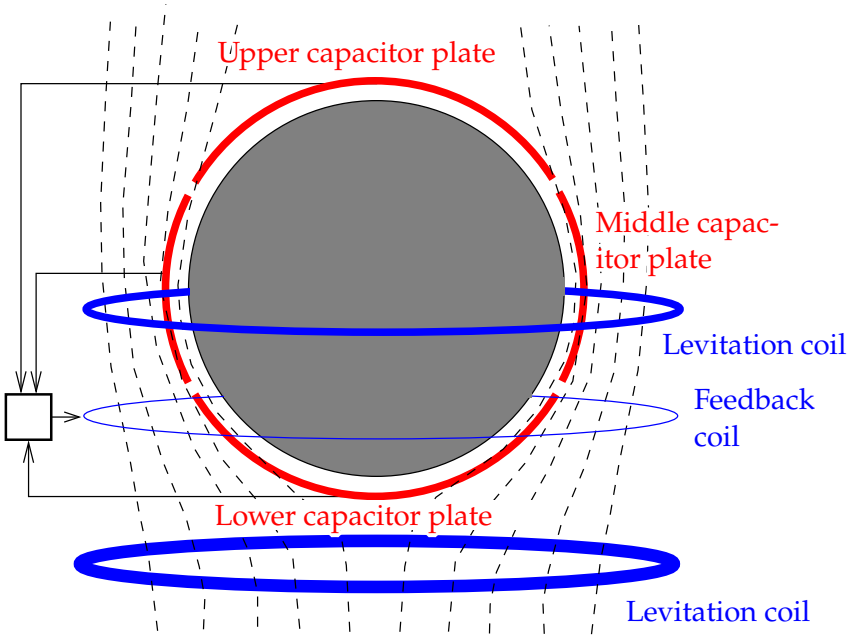


FIGURE 11.9. Operating principle of a superconducting gravimeter. Reading out the sphere position is done capacitively.

The most important property of a superconducting gravimeter is, in addition to its superior precision,[12] its *stability*, the absence of any drift. For this reason, it is extremely suited to monitoring long-period phenomena, like the free oscillations of the solid Earth after large earthquakes,[13] in which the whole Earth tolls like a church bell. Thus it is suitable for measurements that are unsuitable for an ordinary gravimeter because of its larger drift and poorer sensitivity, and measurements for which a seismometer is unsuited because the frequencies are too low.

A recent trend is the development of lightweight, "portable", and

[12] Virtanen (2006) reports how the instrument at Metsähovi detected the change in gravity as workers cleared snow from its laboratory roof, including a tea break! "Weighing" visitors to the lab by their gravitational attraction is also standard fare.

[13] Their periods are in the range of about 300–30 000 seconds — frequencies 0.03–3 mHz — and they are of considerable geophysical interest, Wikipedia, Earth normal modes.

remotely controllable superconducting gravimeters, for example the GWR *i*Grav®, weighing 30 kg and not expending any liquid helium at all. On the other hand it needs over a kilowatt in grid power for its refrigeration system (GWR Instruments, Inc., *i*GRAV® Gravity Sensors). Perhaps this will lead to improvement over the current situation where the bulk of the instruments are located in Europe and North America.

11.6 Gravity measurement and the atmosphere

The atmosphere has the following effects on gravity:

- *Instrumental effects.* These are due to the way the gravimeter is constructed. Putting the instrument in a pressure chamber makes at least the pressure effect go away. In practice it is easier to *calibrate* the instrument in the laboratory and calculate a correction to the field measurements based on the calibration.
- *Attraction of the atmosphere*. This is real gravitation. It contains irregular variations with place and time that we need to remove from the observed gravity values.

  The effect of the atmosphere can be evaluated using a Bouguer-plate approximation: if the air pressure is p, then the surface mass density of the atmosphere is

  $$\kappa = p/\gamma\,,$$

  where γ is a representative gravity value inside the atmosphere. We do not make a large error by using the sea-level value $\gamma \approx 9.81\,\mathrm{m/s^2}$. The standard value of air pressure at sea level is 1013.25 hPa, giving us on sea level[14] $\kappa \approx 10\,329\,\mathrm{kg/m^2}$. The effect of the Bouguer plate is

  $$2\pi G\kappa = 0.43\,\mathrm{mGal} \tag{11.7}$$

[14] So yes, the force acting on a standard 14-inch laptop screen (aspect ratio 16 : 9) is 547 kg… but it doesn't matter as it is not an old-fashioned vacuum cathode-ray tube.

in the upwards direction.

It would however be wrong to apply this value as a correction! The standard atmosphere is in reality a spherical shell inside which the measurements are made, and inside the shell its attraction vanishes, see section 1.4.

Instead, local *variations* in air pressure have a proportional effect. If the air pressure disturbance is $\Delta p = p - p_0$, in which p_0 is mean air pressure, the correction to be made to a gravity measurement will be

$$\delta g_A = 0.43 \frac{\Delta p}{p_0} \text{mGal}.$$

During the passage of a storm or weather front, this beautiful theory collapses, and simple equations give misleading results. Then it is best to just not do any gravity measurements!

- *Including the atmosphere in the mass of the Earth.* This *is not* a correction to be applied to gravity measurements. It is a *reduction* which is applied in the calculation of gravity *anomalies*, if one wants anomalies in which the effect of the atmosphere does not produce a bias.

  Remember that the reference or normal gravity field of GRS80 is defined in such a way that the parameter $GM_\oplus$ refers to the entire mass of the Earth, including the atmosphere, the Earth's attraction as satellites are sensing it (Heikkinen, 1981).

  Therefore, if one wishes to calculate gravity anomalies that have a global mean of zero, one should also reduce measured gravity by computationally moving *the whole atmosphere above the point of measurement* to below the measurement point, for example to sea level.

  The total mass of the atmosphere is

$$M_A = 4\pi\kappa R^2 = 4\pi \frac{p}{\gamma} R^2.$$

  According to Newton its attraction is

$$\frac{GM_A}{R^2} = \frac{4\pi G p}{\gamma},$$

twice the Bouguer-plate atmospheric reduction 11.7 calculated above. This is the value that should be added to the measured gravity values.

One may also think of this value as the change in gravity if the local atmospheric Bouguer plate were *condensed*, Helmert condensation style, to below the measurement location, producing a double planar Bouguer correction.

At sea level, the correction is 0.87 mGal. At height, the correction is

$$0.87\frac{p(H)}{p_0}\text{mGal},$$

in which $p(H)$ and p_0 are the air pressures at height H and at sea level, respectively.

11.7 Airborne gravimetry and GNSS

In the early years of the 1990s GPS, the Global Positioning System, and more generally, satellite positioning, changed *airborne gravimetry* from a difficult technology to something completely operational. To understand this, one must know the principle of operation of airborne gravimetry.

ilma-gravimetria

An aircraft carries an airborne gravimeter, an instrument that, in the same way as a sea gravimeter, is strongly *damped*. The measurement is done automatically, generally using electrostatic compensation. The instrument is mounted on a stabilised platform that follows the local vertical.

During flight, the gravimeter measures *total gravity* on-board the aircraft, consisting of two parts:

1. gravity proper — gravity as sensed in a reference frame connected to the solid Earth and rotating with it
2. the pseudo-forces caused by the inevitable accelerations of the aircraft even in cruise flight.

vertauskehys

Attached to the aircraft are a number of GNSS antennas. With these and a geodetic GNSS instrument, the motions of the aircraft can be monitored with centimetre accuracy. From these motions, we may then calculate the pseudo-forces mentioned above under item 2.

If we measure the position of the plane (or instrument) $\mathbf{x}_i$ at moments t_i, $\Delta t = t_{i+1} - t_i$, we obtain estimated acceleration values as follows (in an inertial frame):

$$\mathbf{a}_i^* \approx \frac{\mathbf{x}_{i+1}^* + \mathbf{x}_{i-1}^* - 2\mathbf{x}_i^*}{\Delta t^2}. \tag{11.8}$$

When the acceleration measured by the gravimeter is $\widetilde{g}$ and the direction
luotiviiva of the local plumb line (upwards) $\mathbf{n}$, local gravity g follows:

$$g = \widetilde{g} - \left\langle \left(\mathbf{a}^* + \mathbf{f}_\omega\right) \cdot \mathbf{n} \right\rangle = \widetilde{g} - \left\langle \mathbf{a}^* \cdot \mathbf{n} \right\rangle - \omega_\oplus^2 \, N(\varphi) \cos\varphi,$$

in which $\mathbf{f}_\omega = \omega_\oplus^2 \left(X\,\mathbf{i} + Y\,\mathbf{j}\right)$ is the centrifugal acceleration of Earth rotation, equation 4.1. $N(\varphi)$ is the transversal radius of curvature of the Earth ellipsoid, equation 2.6.

If working in a co-rotating (Earth centred, Earth fixed) frame, the centrifugal acceleration term should be omitted. However, then a Coriolis acceleration due to the interaction of aircraft velocity $\boldsymbol{v}$ with Earth rotation needs to be introduced instead. This acceleration term is

$$\mathbf{f}'_\omega = -2\left\langle \boldsymbol{\omega}_\oplus \times \boldsymbol{v} \right\rangle = 2\omega_\oplus \left(v_Y\,\mathbf{i} - v_X\,\mathbf{j}\right),$$

and we obtain

$$g = \widetilde{g} - \left\langle \left(\mathbf{a}^\oplus - \mathbf{f}'_\omega\right) \cdot \mathbf{n} \right\rangle = \widetilde{g} - \left\langle \mathbf{a}^\oplus \cdot \mathbf{n} \right\rangle + 2\omega_\oplus \, v_{\text{east}} \cos\varphi\,.$$

The choice of the time constant Δt is critical in this method. It is best to choose it as long as possible, as then, the precision of the calculated GNSS accelerations $\mathbf{a}_i$ is as good as possible. The damping of the gravimeter is also chosen in accordance with Δt, and the observations are filtered digitally: all frequencies above the bound Δt^{-1} are removed, because they are almost entirely caused by the motions of the aircraft.

Often the high-frequency part removed from the signal is 10 000 times stronger than the gravity signal we are after! See for example Lu et al. (2017) figure 2.

If the uncertainty (mean error) of one GNSS vertical position *co-ordinate* measurement is σ_z (and the different co-ordinates do not correlate with each other), then according to equation 11.8 the uncertainty of the vertical acceleration is

$$\sigma_a = \frac{\sigma_z\sqrt{6}}{\Delta t^2}.$$

Making the time interval Δt as long as possible without *resolution* suffering requires a low flight speed. Generally a propeller aircraft or even a helicopter is used. Of course the price of the measurement grows with the duration of the flight — a helicopter rotor hour is expensive!

The flight height H is chosen in accordance with resolution Δx:

$$H \sim \Delta x = v\,\Delta t,$$

where v is the flight speed. The separation between adjacent flight lines is chosen similarly.

The first major airborne gravimetry project was probably the Greenland Aerogeophysics Project (Brozena, 1992). In this ambitious American-Danish project in the summers of 1991 and 1992, over 200 000 km was flown, all the time measuring gravity and the magnetic field, and the height of the ice surface using altimetry.

Since then, other large uninhabited areas in the Arctic and Antarctic regions have also been mapped, see Brozena et al. (1996), Brozena and Peters (1994). Already in subsection 9.6.2 we made mention of other large surveys. Activity continues, see Coakley et al. (2013), Kenyon et al. (2012). The method is very suitable for large, uninhabited areas, but also, for example, for sea areas near the coast or inside archipelagos, where ship gravimeters would have difficulty navigating long straight tracks. In 1999, an airborne gravimetry campaign was undertaken over the Baltic Sea, including the Gulf of Finland (Jussi Kääriäinen, personal communication).

In addition to the economic viewpoint, an important advantage of airborne gravimetry is that *homogeneous coverage* by gravimetric data is obtained from a large area. The homogeneity of surface gravimetric data collected over many decades is difficult to guarantee in the same way. Moreover, the effect of the very local terrain, which for surface measurements is a hard-to-remove systematic error source, especially in mountainous terrain (see section 6.3), does not come into play in the same way for airborne gravimetry.

The operating principle of *satellite gravimetry* is similar, see section 13.7. An essential difference is, however, that the instrumentation on the satellite is in a state of weightlessness: $\widetilde{g} = 0$ (in a high orbit, or when using an air drag compensation mechanism), or $\widetilde{g}$ is small and is measured using a sensitive accelerometer (in a low orbit, where air drag is significant).

ilmanvastus
kiihtyvyys-
mittari

The greatest challenge in planning a satellite gravity mission is choosing the flight height. The lowest possible height is some 200 km. At that height, a tankload of propellant is already needed, or the flight will not last long. However, the resolution of the measurements on the Earth's surface is limited: for example, the smallest details in the Earth's gravity field "seen" by the GOCE satellite are 50–100 km in diameter.

ajoaine

11.8 Measuring the gravity gradient

The acceleration of gravity $\mathbf{g}$ is the gradient of the geopotential W. The acceleration of gravity varies itself with place, especially close to masses. We speak of the *gravity-gradient tensor* or Eötvös tensor:

$$\mathcal{M} \overset{\text{def}}{=} \begin{bmatrix} \frac{\partial^2}{\partial x^2} & \frac{\partial^2}{\partial x \partial y} & \frac{\partial^2}{\partial x \partial z} \\ \frac{\partial^2}{\partial y \partial x} & \frac{\partial^2}{\partial y^2} & \frac{\partial^2}{\partial y \partial z} \\ \frac{\partial^2}{\partial z \partial x} & \frac{\partial^2}{\partial z \partial y} & \frac{\partial^2}{\partial z^2} \end{bmatrix} W = \begin{bmatrix} \partial_{xx} & \partial_{xy} & \partial_{xz} \\ \partial_{yx} & \partial_{yy} & \partial_{yz} \\ \partial_{zx} & \partial_{zy} & \partial_{zz} \end{bmatrix} W.$$

We know that gravity increases going down, at least in free air. Going

up, gravity diminishes, about 0.3 mGal for every metre of height.

In topocentric co-ordinates (x, y, z), where z points to the zenith, this matrix is approximately

$$\mathcal{M} \approx \begin{bmatrix} -0.15 & 0 & 0 \\ 0 & -0.15 & 0 \\ 0 & 0 & 0.3 \end{bmatrix} \text{mGal}/\text{m},$$

in which $\partial_{zz} W = \partial_z g_z = -\partial_z g \approx 0.3\,\text{mGal}/\text{m}$ is the standard value for the vertical free-air gravity gradient: Newton's law gives for a spherical Earth

painovoiman ilmagradientti

$$g_z = -\frac{GM}{(R+z)^2}.$$

The minus sign is because $\mathbf{g}$ points downwards while the z co-ordinate increases going up. Differentiation gives

$$\frac{\partial}{\partial z} g_z = 2\frac{GM}{(R+z)^3} \cdot \frac{\partial\,(R+z)}{\partial z} = -\frac{2g_z}{(R+z)} \approx$$
$$\approx 3 \cdot 10^{-6}\ \text{m}/\text{s}^2/\text{m} = 0.3\,\text{mGal}/\text{m}.$$

The quantities $\partial_{xx} W$ and $\partial_{yy} W$ again represent the curvatures of the equipotential or level surfaces in the x and y directions, equations 4.4:

$$\partial_{xx} W = \frac{\partial^2 W}{\partial x^2} = -\frac{g}{\rho_x}, \quad \partial_{yy} W = \frac{\partial^2 W}{\partial y^2} = -\frac{g}{\rho_y},$$

in which ρ_x and ρ_y are the radii of curvature in the x and y directions. The substitution $\rho_x, \rho_y \approx R$ yields

$$\partial_{xx} W = \partial_{yy} W \approx -1.5 \cdot 10^{-6}\ \text{m}/\text{s}^2/\text{m} = -0.15\,\text{mGal}/\text{m}.$$

The Hungarian researcher Loránd Eötvös did a number of clever experiments (Eötvös, 1998) in order to measure components of the gravity-gradient tensor with *torsion balances* built by him. The method continues to be in use in geophysical research, because the gravity gradient as a measured quantity is very sensitive to local variations in matter density in the Earth's crust.

torsiovaaka, kiertoheiluri

In honour of Eötvös, we use as the unit of gravity gradient the eötvös, symbol E:

$$1\,\mathrm{E} = 10^{-9}\,{}^{\mathrm{m}/\mathrm{s}^2}/_{\mathrm{m}} = 10^{-4}\,{}^{\mathrm{mGal}}/_{\mathrm{m}}.$$

The above tensor is now

$$\mathcal{M} \approx \begin{bmatrix} -1500 & 0 & 0 \\ 0 & -1500 & 0 \\ 0 & 0 & 3000 \end{bmatrix} \mathrm{E}.$$

Note that

$$\frac{\partial^2 W}{\partial x^2} + \frac{\partial^2 W}{\partial y^2} + \frac{\partial^2 W}{\partial z^2} = \partial_{xx} W + \partial_{yy} W + \partial_{zz} W \approx 0,$$

the familiar Laplace differential equation. However, the equation is not exact here: in a co-ordinate system co-rotating with the Earth, the term for the centrifugal force divergence, $2\omega_{\oplus}^2$, must be added, equation 4.2.

The gravity-gradient field of Sun and Moon is known on the Earth's surface as the *tidal field*, see section 14.1.

Self-test questions

1. For the spring gravimeter described in section 11.2, one milligal of change in gravity g produces according to equation 11.4 a lengthening of $5 \cdot 10^{-8}$ m. Do a calculational check.
2. Why is a pendulum gravimeter, although theoretically absolute, not very accurate as an absolute gravimeter?
3. By which method choices do we, in practical measurements, take the drift of a relative gravimeter into account?
4. Why were, before the advent of absolute gravimeters, the reference points of international fundamental gravimetric networks often on airports?
5. In an absolute or ballistic gravimeter, what is the role of:

(a) the "cage" surrounding the falling prism

(b) the "superspring"?

6. According to Google
   - The Gulf War from 1990 to 1991 was the first conflict in which the military used GPS widely.
   - By December 1993, GPS achieved initial operational capability (IOC), indicating a full constellation (24 satellites) being available.
   - The Greenland Aerogeophysics Project, the first ever large-scale airborne gravimetric mission, mapped the gravity field of Greenland during the summers of 1991 and 1992.

   Why are these three dates so close together?

Exercise 11–1: Absolute gravimeter

The observation process of absolute gravimetry is described by the equation

$$z = z_0 + v_0 t + \tfrac{1}{2} g t^2.$$

Assume that the distance of falling is 30 cm.

1. How long is the time of falling?
2. If we aim at an accuracy of $\pm 10\,\mu$Gal, how accurately should the laser interferometer measure the falling distance of the prism?

   Feel free to choose the method of analysis to be used: analytical, numerical, ... Consider that you are in a purchasing situation when building an absolute gravimeter. A ballpark estimate is good enough!
3. The same question for the measurement accuracy of the falling time.

Exercise 11–2: Spring gravimeter

When we use a spring gravimeter in the field, we place the device at every measurement station on a solid base, for example bedrock, for measurement, and *level* it.

Furthermore we always take care that

- The device is *arrested* during transport: the beam is clamped to be motionless.
- The internal temperature of the device is kept constant by a *thermostat system*.

The reason for this is that the functioning of a spring gravimeter depends on the properties of the spring material, which may change as a result of careless handling or temperature variations.

Furthermore, a spring gravimeter always has a *drift*: the connection between measured value and true value changes slowly over time. In a non-factory-fresh gravimeter, this drift is however very regular and almost linear.

Question How is the behaviour of a spring gravimeter, especially its drift, taken into account

1. in planning the topology of the measurement network?
2. in planning the time order of the different measurements in a network?
3. in the choice of vehicles and point locations?

Exercise 11–3: Air pressure and gravity

1. How much does a low-pressure zone of 100 hPa — meaning that the air pressure is 100 hPa *lower* than average air pressure 1013.25 hPa — affect gravity measured on the Earth's surface? Assume the low-pressure zone to be of great areal extent, as low-pressure zones are.

2. How much does sea water rise due to the "inverted barometer effect" under a low-pressure zone? ylösalainen ilmapuntari
3. To how much does the effect from point 2 amount in local gravity measured *on a ship*? Assume that you are on the open sea, that the free-air vertical gravity gradient is $-0.3\,{}^{\mathrm{mGal}}/_{\mathrm{m}}$, and that the density of sea water is $1030\,{}^{\mathrm{kg}}/_{\mathrm{m}^3}$. Analyse the situation carefully.[15] [15]

[15] Meaning *really* carefully.

The geoid, mean sea level, and sea-surface topography

12

12.1 Basic concepts

On the ocean, the geoid is on average at the same level as the *mean sea level*, the surface obtained by removing all periodic and quasi-periodic variations from the instantaneous sea surface. These variations include

- tidal phenomena, caused by Sun and Moon, of an order of magnitude of ± 1 m, locally even more
- variations caused by air pressure variations ("inverted barometer effect", IB). Typically of an order of decimetres, under tropical cyclones up to metres — ylösalainen ilmapuntari
- "wind pile-up", water being pushed by winds
- in littoral seas, variation in the volume of fresh water flowing out from rivers into the sea — makea vesi
- eddies that are formed in the oceans in connection with, for example, the Gulf Stream and the Agulhas Stream ("mesoscale eddies") that may live for months, and inside of which the sea surface may be even decimetres above or below that of the surroundings
- the continual shifting of ocean currents from place to place
- the ENSO, El Niño Southern Oscillation, is a very long time-scale, quasi-periodic weather phenomenon happening in the waters of the Pacific Ocean and the air above it, but affecting weather

phenomena worldwide. The time-scale of variability ranges from two to seven years. See figure 13.1.

If we remove all these periodic and quasi-periodic variations, we are left with the *mean sea level*. If the water of the seas were in a state of equilibrium, this mean sea surface would be an equipotential or level surface of the Earth's gravity field, the *geoid*.

This is, however, not how things really are. Mean sea level differs from a level surface due to for example the following phenomena:

- Permanent ocean currents cause, through the Coriolis force, permanent differences in mean water level.
- suolaisuus Permanent differences in temperature and salinity also cause permanent differences in the mean water level, the latter for example in front of the mouths of rivers.

These physical phenomena, among other things, cause the so-called meritopografia *sea-surface topography*, the permanent separation between the mean sea surface and the geoid. See figure 12.4.

The classical definition of the geoid is

> *"The level surface of the Earth's gravity field that agrees on average best with the mean sea level."*

The practical problem with this definition is that determining the correct level of the geoid requires knowledge of the mean sea level everywhere on the world ocean. This is why many "geoid" models in practice do not coincide with global mean sea level, but with some locally defined mean sea level — and often only approximately.

Mean sea level in its turn is also a problematic concept. It is sea level from which all periodic effects have been computationally removed — but who can know if a so-called secular effect in reality is perhaps long-period? The measure of permanency are the time series that are mareografi available, as tide gauges have been widely operating already for about

a century, when again modern satellite time series — TOPEX/Poseidon and its successors — are just about a quarter of a century long.

A sensible compromise may be the average sea level over 18 years, an important periodicity, *saros* (Wikipedia, Saros), in the orbital motion of the Moon.

12.2 Geoid models and national height datums

A locally determined geoid model is generally *relative*. Locally, at the current state of the art, we have no access to global mean sea level at an acceptable precision. This is likely to change with technology development.

In general, a local geoid model is tied to a *national height system*, and the difference from the classical definition is thus the same as the difference of the national height system from the global mean sea level.

In *Finland*, heights were determined for a long time in the N60 height system, which is tied to the mean sea level in Helsinki harbour at the start of 1960. The difference between it and the global mean sea level is about 30 cm, due to the sea-surface topography in the Baltic Sea, see figure 12.4. The *reference benchmark* of the system is located in nearby Kaivopuisto, figure 7.2. Precise levelling disseminated heights from here all over Finland.

pääkiintopiste

The modern Finnish height system is N2000, which is in principle tied to the mean sea level in Amsterdam, which is close to global mean sea level. The reference benchmark in Finland is similarly located at the Metsähovi research station in the Kirkkonummi municipality, west of Helsinki.

At the beginning of 1960, the reference surface of the Finnish height system N60 was an equipotential or level surface of the Earth's gravity field. However, due to uplift, that is no longer the case: the *post-glacial land uplift* varies from some four millimetres per year in the Helsinki area to ten millimetres per year in the area of maximum land uplift near

Pohjanmaa

Ostrobothnia. This is the main reason why in Fennoscandia, height systems have a "best before" date and must be modernised a couple of times per century.

Generally, geoid maps for practical use, like FIN2000, the Finnish geoid model (figure 9.5), are constructed so that they transform heights in the national height system, for example N60 heights (Helmert heights) above "mean sea level" to heights above the reference ellipsoid of the GRS80 system.

vertaus-ellipsoidi

As land uplift is an ongoing process, it must be tied to a certain epoch, a point in time at which the GNSS measurements were done to which the original gravimetric geoid solution has been fitted. In the case of FIN2000 this was 1997.0 (Matti Ollikainen, personal communication; Bilker-Koivula and Ollikainen, 2009; Häkli et al., 2009).

Strictly speaking then, FIN2000 is not a model of the geoid. A better name might be "transformation surface". This holds true, in fact, for *all* national or regional geoid models that are built primarily for the purpose of enabling the use of GNSS in height determination ("GNSS levelling"). These "geoid-like surfaces" are generally constructed in the following way:

poistamis-entistämis-menetelmä

1. Calculate a gravimetric geoid model by using the Stokes integral equation and remove–restore, for example using the FFT calculation technique.
2. Fit this geoid surface solution to a number of comparison points, in which both the height from levelling — "above sea level" — and from the GNSS method — above the reference ellipsoid — are known. The fit takes place for example by modelling the differences as a polynomial function:

$$\delta N = a + b\,(\lambda - \lambda_0) + c\,(\varphi - \varphi_0) + \cdots$$

or something more complicated, and solving the coefficients a, b, c, ... from the differences between the two heights in these known comparison points by using the least-squares method.

pienimmän neliösumman menetelmä

12.3 The geoid and post-glacial land uplift

Global mean sea level is not constant. It rises slowly by an amount that, over the past century, has slowly grown. Over the whole 20th century, the rate has been 1.5–2.0 mm/a, for example 1.6 mm/a (Wöppelmann et al., 2009). Over the last several decades, the rate has accelerated and is now over 3 mm/a, see figure 13.1.

This value is called the *eustatic rise of mean sea level*. It is caused partly by the melting of glaciers, ice caps, and continental ice sheets, partly by the thermal expansion of sea water. A precise value for the eustatic rise is hard to determine: almost all tide gauges used for monitoring sea level have their own vertical motions, and distinguishing these from the rise in sea level requires a representative geographic distribution of measurement locations. In particular the ongoing isostatic response of the solid Earth to the ending of the last ice age, the latest deglaciation or termination, the so-called GIA (glacial isostatic adjustment), is a global phenomenon that it has only been possible to observe by satellite positioning in the most recent decades.

mannerjäätikkö

mareografi

Because of eustatic sea-level rise, a distinction must be made between *absolute* and *relative* land uplift:

Absolute land uplift is the motion of the Earth's crust relative to the centre of mass of the Earth — or equivalently, relative to the surface of a geocentric reference ellipsoid such as GRS80. This land uplift is measured when using satellites the orbits of which are determined in a co-ordinate reference system tied to the Earth's centre of mass, for example, positioning of tide gauges by means of GNSS.

vertaus-järjestelmä

Relative land uplift is the motion of the Earth's crust relative to the mean sea level. This motion is measured by tide gauges, also called *mareographs*.

Geoid rise As the post-glacial land uplift is the shifting of masses internal to the Earth from one place to another, it is clear that

the geoid must also change. The geoid rise is, however, small compared to the land uplift, only a few percent of it.

Equation (the dot above a quantity denotes the derivative with respect to time[1]):

$$\dot{H} = \dot{h} - \dot{N} = \dot{H}_r + \dot{H}_e + \dot{H}_t,$$

in which

$\dot{H}$	relative land uplift from the geoid
$\dot{h}$	absolute land uplift from the reference ellipsoid
$\dot{H}_r$	relative land uplift from the local mean sea level
$\dot{H}_e$	eustatic (global mean sea level) rise
$\dot{H}_t$	change over time of the sea-surface topography (likely small)
$\dot{N}$	geoid rise from the reference ellipsoid.

The rise of the geoid as a result of land uplift can be simply calculated with the Stokes integral equation:

$$\frac{dN}{dt} = \frac{R}{4\pi\gamma} \iint_\sigma S(\psi) \left(\frac{d}{dt}\Delta g\right) d\sigma.$$

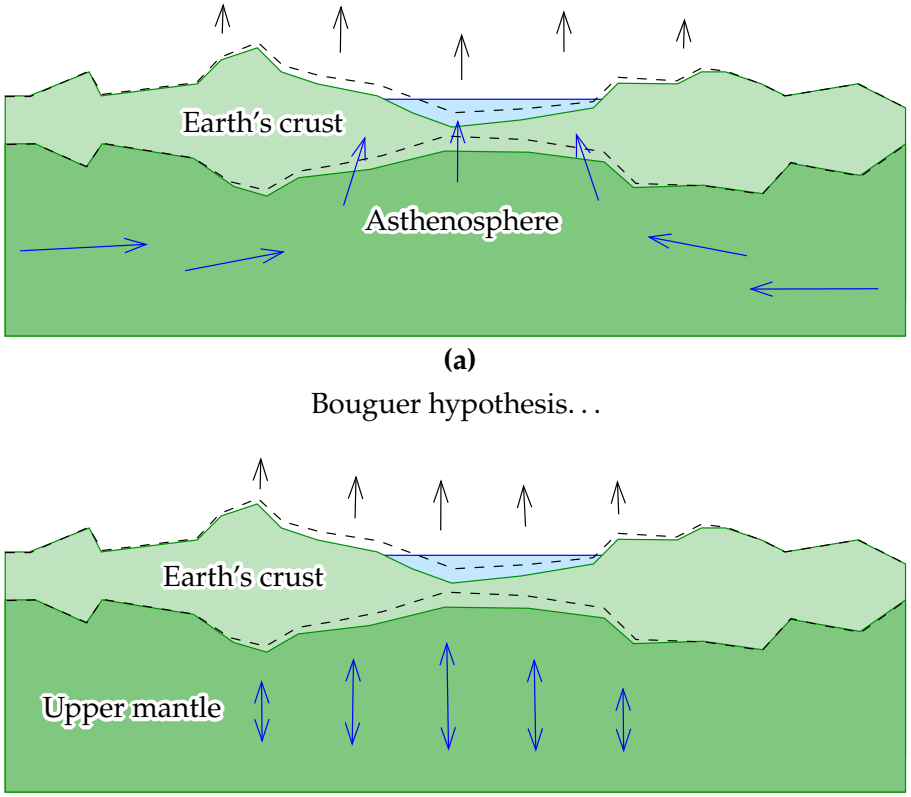
Here, $\frac{d}{dt}\Delta g$ is the change of gravity anomalies over time due to land uplift. Unfortunately we do not precisely know the *mechanism* by which mass flows in the Earth's mantle to underneath the land-uplift area. We may posit

$$\frac{d}{dt}\Delta g = c\frac{dH}{dt} = c\dot{H},$$

in which the constant c may range from -0.16 to $-0.31\,{}^{\text{mGal}}/_{\text{m}}$.

- The value $-0.16\,{}^{\text{mGal}}/_{\text{m}}$ is called the "Bouguer hypothesis": it corresponds to the situation in which upper mantle matter flows into the space freed up underneath the rising Earth's crust, filling it. This matter may be roughly modelled as a Bouguer plate.

[1] This dot notation, fluxion, was introduced by Newton.



(a)

Bouguer hypothesis. . .

(b)

. . . and free-air hypothesis.

FIGURE 12.1. The two different hypotheses on the mechanism of post-glacial land uplift.

- The value $-0.31\,^{\mathrm{mGal}}/_{\mathrm{m}}$ is the opposite extreme, the "free-air hypothesis". By this hypothesis, the ice load during the last ice age has only compressed the Earth's mantle, and now it is slowly expanding again to its former volume (the "rising dough model"). pullataikina-malli

Up until fairly recently, the most likely value was about $-0.2\,^{\mathrm{mGal}}/_{\mathrm{m}}$, with substantial uncertainty. The latest results (Mäkinen et al., 2010; Olsson et al., 2019) may be summarized as $-0.16 \pm 0.02\,^{\mathrm{mGal}}/_{\mathrm{m}}$ (one standard deviation), which would seem to settle the issue. It looks like the Bouguer hypothesis is closer to physical reality. The flow of mass is

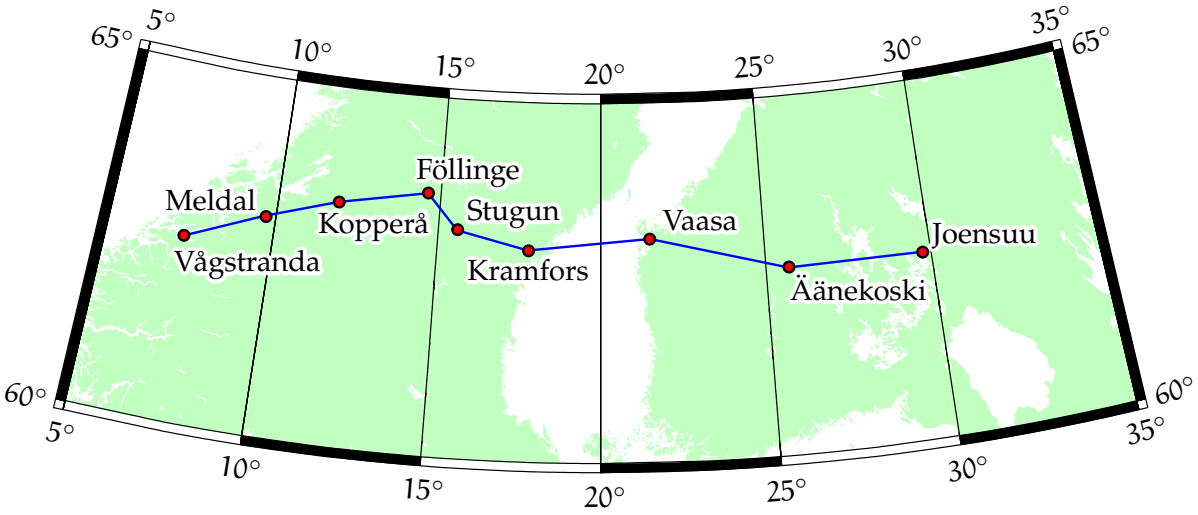


FIGURE 12.2. The Fennoscandian gravity line on the 63rd parallel north.

assumed to happen within the *asthenosphere*.

This problem has been studied much in the Nordic countries. The method used has been gravimetric measurement along the 63rd parallel north (the "Blue Road Geotraverse" project). The measurement stations extend from the Norwegian coast to the Russian border, and have been chosen such that the gravity along them varies within a narrow range. In this way, the effect of the scale error of the gravimeters is avoided. Clearly, absolute gravity is of no interest here, only the *change* in gravity *differences* over time between the stations.

These measurements have been made over many years using high-precision spring or relative gravimeters. In recent years, there has been a shift to using absolute gravimeters, obviating the need for measurement *lines*.

12.4 Determining the sea-surface topography

In principle three geodetic methods exist:

- satellite radar altimetry and gravimetric geoid determination

- positioning of tide gauges along the coast using GNSS, together with gravimetric geoid determination
- precise levelling along the coast connecting tide gauges.

In addition to this, we still have the oceanographic method: physical modelling. The method is termed *steric levelling* if temperature and salinity measurements along vertical profiles are used on the open ocean, and *geostrophic levelling* if ocean current measurements are used to determine the Coriolis effect, generally close to the coast.

All methods should give the same results. The Baltic Sea is a textbook example where all three geodetic methods have been used. It has been found that the whole Baltic Sea surface is tilted: relative to a level surface, the sea surface goes up from the Danish straits to the bottoms of the Gulf of Finland and the Bothnian Bay by 25–30 cm. pohjukka

Oceanographic model calculations show that this tilt is mainly due to a *salinity gradient*: in the Atlantic Ocean, the salinity is 30–35 ‰, when in the Baltic it drops to 5–10 ‰, due to the massive production of fresh water by the rivers (Ekman, 1992). Of course on top of this come temporal variations, like oscillations caused by storms resembling those in a bathtub, the amplitude of which can be over a metre.

In Ekman (1992) more is said about the sea-surface topography of the Baltic and its determination.

12.5 Global sea-surface topography and heat transport

One important reason why researchers are interested in the global sea-surface topography is that it offers an opportunity to study more precisely the currents in the oceans and thus the transport of the Sun's energy from the equator to higher latitudes. There are many other things that a better knowledge of sea currents would help to explore: for example carbon dioxide dissolved into the water, chlorophyll (phytoplankton), and salinity.

The Coriolis force, or *acceleration*, caused by the Earth's rotation is

$$\mathbf{f}'_{\omega} = -2\langle \boldsymbol{\omega}_{\oplus} \times \boldsymbol{\nu} \rangle, \tag{12.1}$$

in which $\boldsymbol{\nu}$ is the velocity vector of a freely moving particle in a system attached to the rotating Earth, and $\boldsymbol{\omega}_{\oplus}$ is the rotation vector of the Earth. This is an axial vector, pointing in the direction of the Earth's axis of rotation.

If a fluid flows on the Earth's surface, then, in the above equation 12.1, only the part of $\boldsymbol{\omega}_{\oplus}$ in the normal direction $\mathbf{n}$ to the ocean surface will have an effect: this part amounts to $\langle \boldsymbol{\omega}_{\oplus} \cdot \mathbf{n} \rangle = \omega_{\oplus} \sin \varphi$, or as a vector,

$$\overline{\boldsymbol{\omega}}_{\oplus} \stackrel{\text{def}}{=} \langle \boldsymbol{\omega}_{\oplus} \cdot \mathbf{n} \rangle \, \mathbf{n} = (\omega_{\oplus} \sin \varphi) \, \mathbf{n}.$$

Now, the Coriolis vector projected onto the horizontal plane is

$$\mathbf{f}'_{\overline{\omega}} \stackrel{\text{def}}{=} -2 \left\langle \overline{\boldsymbol{\omega}}_{\oplus} \times \boldsymbol{\nu} \right\rangle = -2\omega_{\oplus} \sin \varphi \left\langle \mathbf{n} \times \boldsymbol{\nu} \right\rangle,$$

the length of which is the scalar

$$f'_{\overline{\omega}} = \|\mathbf{f}'_{\overline{\omega}}\| = 2\nu \, \omega_{\oplus} \sin |\varphi| \, .$$

Here, $\nu = \|\boldsymbol{\nu}\|$ and $\omega_{\oplus} = \|\boldsymbol{\omega}_{\oplus}\|$ in the familiar way. The *direction* of the Coriolis acceleration is always perpendicular to the flow velocity: when watching along the flow direction, to the right in the northern hemisphere, and to the left in the southern hemisphere.

As a result of the Coriolis force, the sea surface in the area of an ocean current is *tilted sideways* with respect to the current, at an angle

$$\frac{f'_{\overline{\omega}}}{\gamma} = 2\nu \frac{\omega_{\oplus}}{\gamma} \sin |\varphi| \, .$$

Here, γ is local gravity. This equilibrium between Coriolis force and gravity is called the *geostrophic equilibrium*. On the equator, it can be seen from the equation that the tilt is zero, but everywhere else, ocean currents are tilted.

For example, in the case of the Gulf Stream, the height changes caused by this effect are several decimetres. If we define a local (x, y) coordinate system in which $x(\varphi, \lambda)$ is pointing north and $y(\varphi, \lambda)$ east, we may write for the sea-surface topography H the geostrophic equations

$$\frac{\partial H}{\partial x} = -2v_y \frac{\omega_\oplus}{\gamma} \sin\varphi, \quad \frac{\partial H}{\partial y} = +2v_x \frac{\omega_\oplus}{\gamma} \sin\varphi. \tag{12.2}$$

As we will see in chapter 13, we can measure the location in space of the sea surface at a precision of a few centimetres using satellite radar altimetry. If we furthermore have a precise geoid map, we may calculate the sea-surface topography, and with the aid of equations 12.2 solve for the flow velocity vector field[2]

$$\begin{bmatrix} v_x(x,y) & v_y(x,y) \end{bmatrix}^T = \begin{bmatrix} v_x(\varphi,\lambda) & v_y(\varphi,\lambda) \end{bmatrix}^T.$$

An elegant property of these equations is that we do not even have to know the absolute level of the field $H(x, y) = H(\varphi, \lambda)$, because that vanishes in differentiation.

The method described, figure 12.3, requires a sufficiently precise geoid map of the world ocean. The GOCE satellite fits this need like a glove, see subsection 13.7.3. One objective of the mission was, as the name indicates, to get a full picture of ocean currents and especially their capacity for heat transport. This knowledge helps understand how the Earth's climate functions and how it is changing, also as a result of human activity. This is an important issue for Europe and Fennoscandia, and also Finland, as the heat energy brought by the Gulf Stream helps to keep these areas habitable (Caesar et al., 2018).

Even without a geoid model, we can study the *variations* of ocean currents using satellite altimetry. It has long been known that in the

[2] A popular, though unofficial, unit for ocean current is the sverdrup (Wikipedia, Sverdrup), a million cubic metres per second. All the rivers of the world together make about one sverdrup, while the Gulf Stream is 30–150 Sv. "There is a river in the ocean" – Matthew Fontaine Maury (1806–1873), American polymath and pioneer of oceanography.

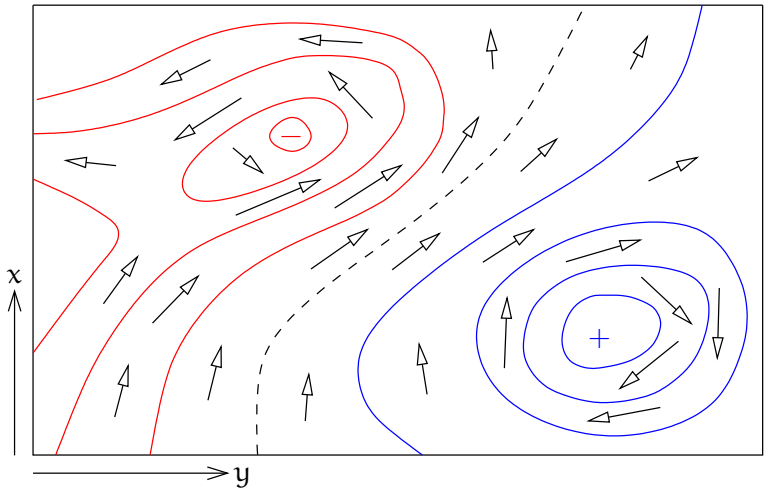


FIGURE 12.3. Connection between sea-surface topography and ocean currents. Arrows depict ocean currents, curves, sea-surface topography.

North Atlantic Ocean, *mesoscale eddies* have been moving alongside the Gulf Stream; eddies of size 10–100 km which show up in altimetric imagery. It is interesting that the eddies also show up in maps of the ocean surface temperature, and biologists have observed that life inside the eddies differs from that outside them (Godø et al., 2012). The life span of the eddies can be weeks, even months.

A good, though somewhat dated, introduction to "geodetic oceanography" is given by Rummel and Sansó (1992).

12.6 The global behaviour of the sea level

Water exists on the Earth in three phases: liquid, ice, and vapour. During geological history, the ratio between liquid water and ice in particular has varied substantially. Also today, a large amount of ice is tied up in continental ice sheets, specifically in Antarctica and Greenland. Of these, the Eastern Antarctic ice sheet is overwhelmingly the largest.

When the amount of water tied up in continental ice sheets varies, so does the sea level. The end of the last ice age has raised the mean

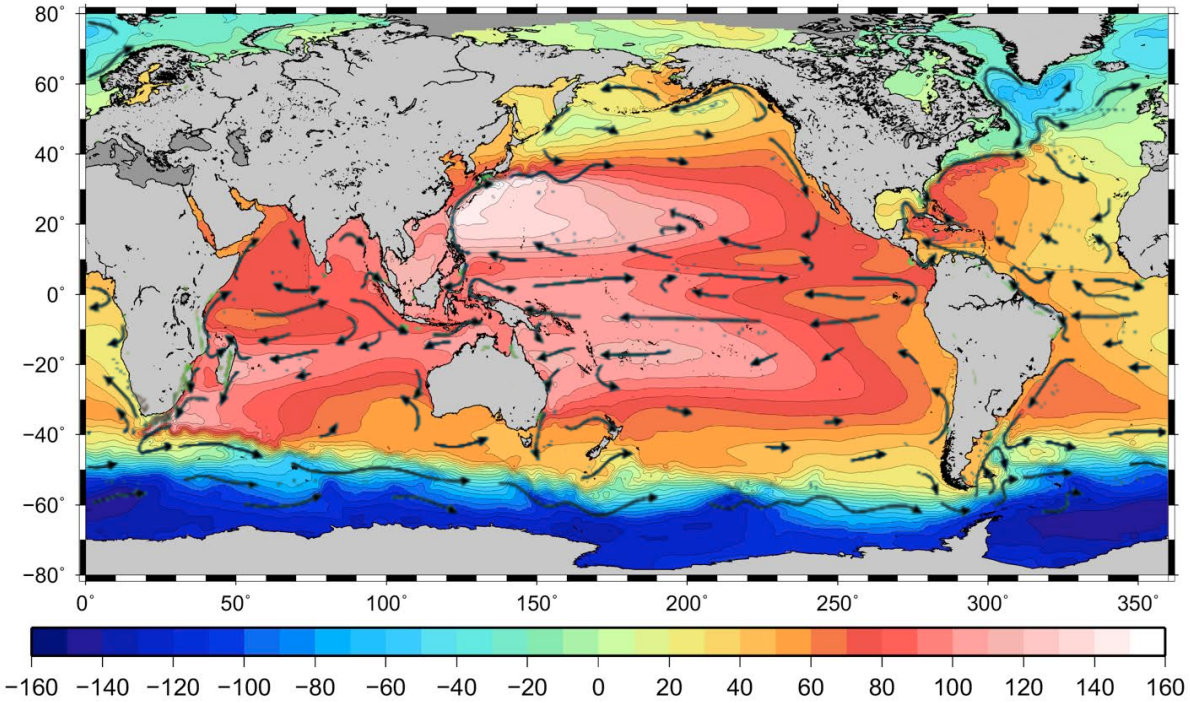


FIGURE 12.4. Sea-surface topography map produced by GOCE. Base map © European Space Agency (ESA). Unit: cm. Ocean currents superimposed: NOAA / Rick Lumpkin (NOAA, Ocean currents).

sea level by as much as 120 m, a process that ran to completion some 7000 years ago[3] (Wikipedia, Sea level rise). Not until the last century or two has the sea level again started rising, and the rise accelerating, as a consequence of global warming.

We still live in the aftermath of the last glaciation. There were large continental ice sheets which have since melted away, like in Fennoscandia and in Canada (the Laurentide ice sheet): the land is still rising at an even pace, up to 10 and 14 millimetres per year, respectively. Around the land-uplift areas, in central Europe and the United States,

[3] 7000 years "before present", 7 ka BP. *BP* conventionally means: before 1950. Nowadays *b2k*, before the year 2000, is also used.

vajoaminen

a *subsidence* of the land is taking place at an annual rate of 0.5–1.7 millimetres, for example DeJong et al. (2015). Directly underneath the hard crust of the Earth or lithosphere, in the upper mantle layer called *asthenosphere*, material is flowing slowly inwards under the rising Earth's crust.

In order to complicate the picture, the sea-level rise caused by the melting of continental ice sheets also presses the ocean floor down — by as much as 0.3 mm per year; the so-called Peltier effect (Peltier, 2009). Therefore, the measured sea-level rise — whether on the coast by tide gauges, or from space using satellite altimetry — *does not represent the whole change in total ocean water volume*. If the latter is what interests us, as it always does in climate research, this Peltier correction must still be added to the observation values.

merenpohjan vajoaminen

The subsidence of the sea floor has not been globally uniform: at the edges of the continents a "lever motion" happens when the sea floor subsides but the dry land does not. And in the tropics in the Indian and Pacific Oceans, the sea level reached its maximum level, the *mid-Holocene highstand*, relative to the Earth's crust approximately 7000 years ago. After this, the local sea level subsided and the coral formations from that age remain, dead, some 2–3 m above the modern sea level. This is how, for example, Tuvalu and the Maldives were formed, which are now being threatened by the modern sea-level rise again.

12.7 The sea-level equation

Scientifically the variations in sea level are studied using the *sea-level equation*. A pioneer in this field has been the Canadian Richard Peltier, who has constructed physics-based models of how both the solid Earth and sea level respond when the total mass of the continental ice sheets changes.

The sea-level equation is (Farrell and Clark, 1976; Spada and Melini,

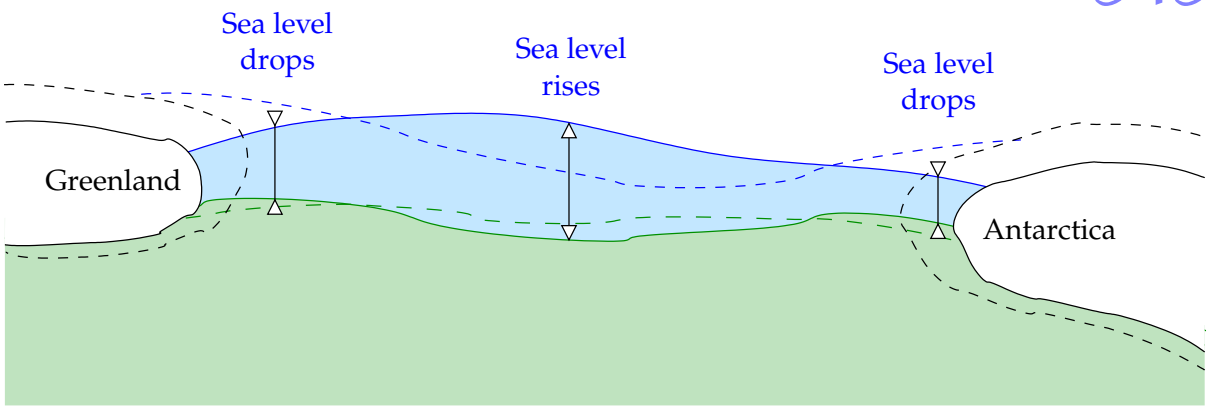


FIGURE 12.5. The sea-level equation. Sea level reacts in a complicated way when continental ice sheets melt.

2015):

$$S = S_E + \frac{G}{R}\Big(\rho_i\big(G_s \otimes_i I - \overline{G_s \otimes_i I}\big) + \rho_w\big(G_s \otimes_o S - \overline{G_s \otimes_o S}\big)\Big), \quad (12.3)$$

in which

- $S = S(\omega, t) = S(\phi, \lambda, t)$ means the variations of sea level as a function of place $\omega = (\phi, \lambda)$ and time t. These variations are relative to the solid Earth's surface: they are changes in sea depth. S is also what tide gauges measure.
- $I = I(\omega, t)$ is similarly a function of place and time describing the variations in thickness of ice sheets and glaciers.
- S_E is the *eustatic term*, the variation in ice volume converted into "equivalent global sea-level variation", in an equation

  $$S_E(t) = -\frac{m_i(t)}{\rho_w A_o},$$

  in which $m_i(t)$ is the variation in total ice mass as a function of time, ρ_w the density of sea water, and A_o the total surface area of the oceans.
- R is the mean radius of the Earth, G Newton's universal gravitational constant, section 1.2.

- ρ is the density of matter: ρ_i that of ice, and ρ_w that of sea water.
- $\otimes$ is the symbol of a convolution over the surface of the Earth and the time axis, $\otimes_i$ over land ice, $\otimes_o$ over the oceans: *Green's function* is multiplied with the ice and sea functions and integrated over the domain in question. These integrals are, by the way, very similar to the ones discussed in section 8.1. For example

$$\begin{aligned}\{G_s \otimes_o S\}(\omega, t) &= \\ &= \int_{-\infty}^{t} \iint_{\text{ocean}} G_s\big(\psi(\omega,\omega'), (t-t')\big)\, S(\omega', t')\, d\omega'\, dt', \end{aligned} \tag{12.4}$$

  in which $\psi(\omega, \omega')$ is the geocentric angular distance between evaluation point $\omega = (\phi, \lambda)$ and data point $\omega' = (\phi', \lambda')$. The measure of the surface integral is $d\omega = R^2\, d\sigma = R^2 \cos\phi\, d\phi\, d\lambda$. As can be seen, we have here a convolution applied *both* over the Earth's surface ω *and* over the time axis t.
- The *overbar* designates *averaging* over the whole relevant surface area.
- G_s is the *Green's function of sea level*

$$G_s(\psi, \Delta t) = \frac{1}{\gamma} G_V(\psi, \Delta t) - G_r(\psi, \Delta t), \tag{12.5}$$

  in which $\Delta t \overset{\text{def}}{=} t - t' \geqslant 0$.

  This equation expresses simply that the sea depth S is the distance between sea surface and sea floor, and that a change in it is the difference between the vertical displacements of those two: that of the sea surface caused by a change in potential V, and that of the sea floor, meaning a change in local radius r.

  Here the *Green's function of the geopotential* is

$$G_V(\psi, \Delta t) = G_V^r(\psi, \Delta t) + G_V^e(\psi, \Delta t) + G_V^v(\psi, \Delta t),$$

  The function $G_V^r(\psi, \Delta t) = \delta(\Delta t) \big/ 2\sin\left(\frac{1}{2}\psi\right)$ is the *rigid* partial Green's function, representing the change in potential caused by a mass of water or ice before any deformation takes place.

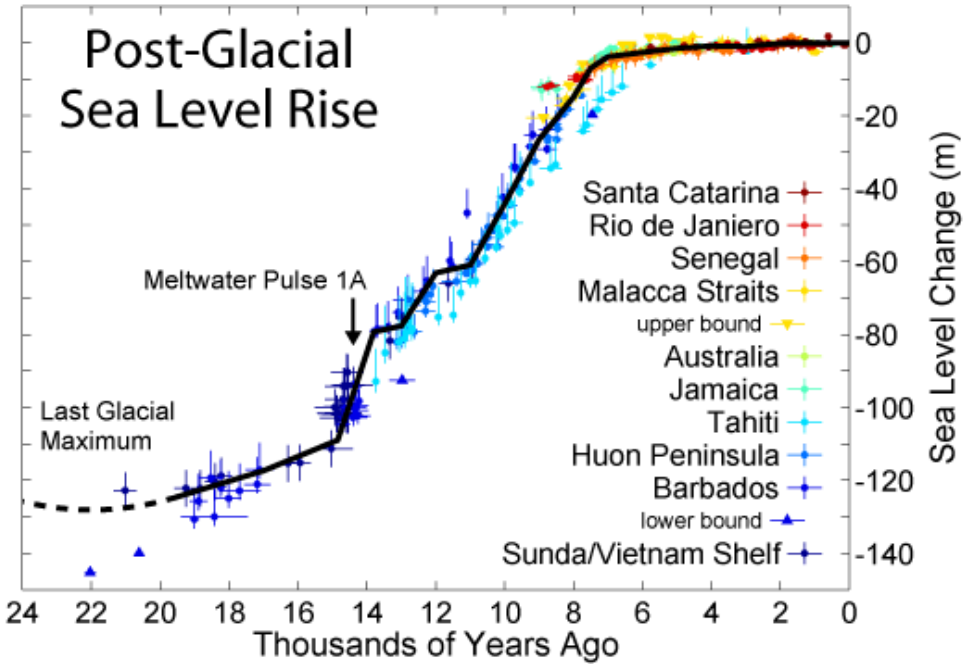


FIGURE 12.6. Sea-level rise after the last ice age (Rohde, 2005).

The functions G^e_V and G^v_V are the elastic and viscous deformation partial Green's functions of the geopotential. These thus characterise the *rheological behaviour* of the Earth, and their theoretical calculation requires the internal density and viscosity distributions $\rho(r)$ and $\eta(r)$ of the Earth — assuming they are *isotropic*, only dependent upon r.

$$G_r(\psi, \Delta t) = G^e_r(\psi, \Delta t) + G^v_r(\psi, \Delta t)$$

is similarly *Green's function of vertical* or *radial displacement* of the sea floor, in the same way split into elastic and viscous parts. There trivially is no "rigid" part.

The behaviour of the sea level can now be computed in this way where one first tries to construct an "ice-load history" $I(\omega, t)$. Then, from this one tries to calculate iteratively, using sea-level equation 12.3, $S(\omega, t)$. S signifies *relative* sea-level variation, changes in the vertical difference

in location between sea level and the Earth's solid body or Earth's crust. It is a *function of place*: one may not assume that it would be the same everywhere. Mitrovica et al. (2001) show how, for example, the meltwater from Greenland flees to the southern hemisphere, when the meltwater from Antarctica again comes similarly to the north. This is a consequence of the change in the *Earth's gravity field* and the geoid, when large volumes of ice melt. And also the physical *shape* of the Earth changes when the ice load changes: *glacial isostatic adjustment*, GIA.

This also complicates the monitoring of the global mean sea level from local measurements: the problem is familiar in Fennoscandia, where the Earth's crust, for now, is rising faster than the global sea level.

Green's functions in the sea-level equation are functions of both ψ and time difference Δt. This tells us that GIA is a function of both place and time. On a spherically symmetric Earth, the functions may be written as *expansions*. See Wieczerkowski et al. (1999).

The *elastic* response of the Earth to loading is instantaneous on the geological time-scale. It is described by similar elastic Love numbers as those that appear in the theory of tidal deformation, for long (though geologically short) periods P. See section 14.2. Like this:

$$G^e_s(\psi,\Delta t) = \frac{1}{\gamma}\cdot \overbrace{\delta(\Delta t)\sum_{n=0}^{\infty} k_n P_n(\cos\psi)}^{G^e_V(\psi,\Delta t)} - \overbrace{\frac{1}{\gamma}\cdot \delta(\Delta t)\sum_{n=0}^{\infty} h_n P_n(\cos\psi)}^{G^e_r(\psi,\Delta t)},$$

with k_n and h_n also appearing in equations 14.4. $\delta(\Delta t)$ is Dirac's delta function.

GIA, however, is *viscous* deformation on a range of geological time-scales. Equation 12.5 becomes

$$G^v_s(\psi,\Delta t) = \frac{1}{\gamma}\cdot \overbrace{\sum_{n=0}^{\infty} k^v_n(\Delta t)\, P_n(\cos\psi)}^{G^v_V(\psi,\Delta t)} - \overbrace{\frac{1}{\gamma}\sum_{n=0}^{\infty} h^v_n(\Delta t)\, P_n(\cos\psi)}^{G^v_r(\psi,\Delta t)},$$

with the *viscous Love numbers* for potential and vertical displacement:

$$k_n^v(\Delta t) = \sum_{i=1}^{I} r_{ni}^k \exp(-s_{ni}\,\Delta t), \quad h_n^v(\Delta t) = \sum_{i=1}^{I} r_{ni}^h \exp(-s_{ni}\,\Delta t).$$

Here, n is the degree number, and the index $i = 1, \ldots, I$ counts the *viscous relaxation modes* for every degree number n. The number of different modes I is in practice a handful, each related to a different discontinuity surface in the Earth's density and viscosity model used. The ratios r_{ni}^k/s_{ni} and r_{ni}^h/s_{ni} are called "modal strengths", and the $\tau_{ni} \stackrel{\text{def}}{=} 1/s_{ni}$ *are relaxation times* in which the mode in question will decay over time.

Generally, uplift patterns that are of large spatial extent — low degree numbers n — decay slower, when again the local patterns — high degree numbers — tend to decay faster. The local uplift patterns of the last deglaciation have today already vanished: the Fennoscandian land uplift is already geographically very smooth and the seismicity accompanying the deglaciation is pretty much over. Back then, during the retreat of the ice sheet at its edge, there were strong earthquakes, traces of which are visible in the landscape (Kuivamäki et al., 1998). The now-dominant viscous uplift patterns are many hundreds of kilometres in geographic extent and correspondingly of time-scales of thousands of years.

Self-test questions

1. List all the causes of sea-level variations that you are aware of.
2. What is the sea-surface topography?
3. What is eustatic sea-level rise?
4. What is the origin of the name "*El Niño*"?
5. What is absolute, and what is relative land uplift? What does the difference between the two consist of?

6. Which two main models are on offer for the mechanism of land uplift?
7. Which three geodetic techniques are available for determining the sea-surface topography?
8. What is the shape of the sea-surface topography of the Baltic Sea, and what is its cause?
9. What is the Coriolis force, and how does it affect ocean currents?
10. What is the geostrophic balance?
11. In whose honour is the unit sverdrup named?
12. How can one invert a map of the sea-surface topography into a map of ocean currents? Where on Earth does this *not* work?
13. What is the Peltier effect? What is the mid-Holocene highstand?
14. What does the sea-level equation describe?
15. How would sea-level equation 12.3 change if the convolution integrals like equation 12.4 were over the unit sphere $\mathrm{d}\sigma = \cos\phi\,\mathrm{d}\phi\,\mathrm{d}\lambda$ instead of over $\mathrm{d}\omega = R^2\cos\phi\,\mathrm{d}\phi\,\mathrm{d}\lambda$?
16. Why does the mean sea level in the Baltic Sea not rise when the Greenland continental ice sheet melts? What will happen in the Baltic Sea when the West Antarctic ice sheet melts?

Exercise 12−1: Coriolis force, ocean current

It is given that the velocity of flow of an ocean current is $0.1\,\mathrm{m/s}$ and its width 100 km.

1. How much is the height difference between its left and right edges? Which edge is higher? Assume that the current is at latitude 45° north.
2. If the same current were 200 km broad and the velocity of flow $0.05\,\mathrm{m/s}$ (so, if the same depth is assumed, the amount of water transported would also be the same), compute for that case the

height difference between the left and the right edges.

3. (*For fun*) if the depth of the current is 1 km, what is the water transport in sverdrup?

Exercise 12–2: Land subsidence and the mechanism of land uplift

How does the post-glacial land subsidence observed in the United States and central Europe support a Bouguer type of land-uplift mechanism (figure 12.1a), rather than a free-air mechanism?

Satellite altimetry and satellite gravity missions

13

13.1 Satellite altimetry

Satellite altimetry is a measurement method in which the distance from a satellite straight down to the sea surface is measured using a microwave radar. Historically there have been many satellites carrying an altimetry radar: see table 13.1, which may not be complete.

- The GEOS-3 (1975-027A) and Seasat satellites were American testing satellites aimed at developing the altimetric technique. The measurement precision of GEOS-3 was still modest. Before that, satellite altimetry was also tested with a device, accuracy ± 1 m, on board the orbital laboratory Skylab (1973-027A).
- Seasat (1978-064A) broke down only three months after launch, probably due to a short-circuit.[1] However, the data from Seasat was the first large satellite altimetry data set used for determining the global mean sea surface, and that of the Baltic Sea (Vermeer, 1983b). [1]
- Geosat (1985-021A) was a satellite launched by the US Navy, intended to map the gravity field on the world ocean, more precisely the deflections of the plumb line, which are needed to impart the correct departure direction to ballistic missiles

luotiviivan poikkeama

[1] But read this: Wikipedia, Seasat conspiracy theory.

TABLE 13.1. Altimetric satellites through the ages.

Satellite	Launch year	Orbital inclination (°)	Orbital height (km)	Repeat periods (days)	Measurement precision (m)	Positioning technique
GEOS-3	1975	115.0	843	~ 38	0.20	
Seasat	1978	108.0	780	3, 17.07	0.08	
Geosat	1985	108.05	786	3, 17.07	0.04	
ERS-1	1991	98.5	780	3, 35, 168	0.03	
TOPEX/Poseidon	1992	66.0	1337	9.9156	0.033	GPS, DORIS
ERS-2	1995	98.5	780	3, 35	0.03	PRARE
Geosat follow-on	1998	108.0	800	17.07	0.035	
Envisat	2001	98.5	784	35	0.045	GPS, DORIS
Jason-1	2001	66.1	1336	9.9156	0.025	GPS, DORIS
Jason-2	2008	66.04	1336	9.9156	0.025	GPS, DORIS
CryoSat-2	2010	92.0	725	369		DORIS
Haiyang-2A	2011	99.3	970	14, 168	0.085	DORIS, GPS
SARAL/AltiKa	2013	98.5	781	35		DORIS
Jason-3	2016	66.04	1338	9.9927	0.025	GPS, DORIS
Sentinel-3A	2016	98.62	804	27	0.03	DORIS, SLR, GNSS

launched from submarines. The 17-day repeat data from the geodetic mission was initially classified. Later however, the data from the southern hemisphere was published for scientists to use, and still later, the whole data set was made public.

- The satellites ERS-1/2 (1991-050A, 1995-021A) and Envisat (2002-009A) were launched by the ESA, the European Space Agency. The altimeter was just one among many packages. A German positioning instrument called PRARE was on the ERS satellites, but it only functioned after launch on ERS-2.
- TOPEX/Poseidon (1992-052A) was an American-French collaboration, one goal of which was to precisely map the *sea-surface topography*. A special feature was the on-board precise GPS positioning device, which allowed the determination of the location

meritopografia

of the sea surface geocentrically. Together with its successors Jason-1, 2 and 3 (2001-055A, 2008-032A, 2016-002A), this satellite mission has also produced, and continues to produce, valuable information on the global rise of the sea level over the last 25 years, of about 3 mm per year. See figure 13.1.

The famous oceanographer Walter Munk[2] characterised TOPEX/Poseidon in 2002 as "the most successful ocean experiment of all time" (Munk, 2002).

- Haiyang-2A (2011-043A) is a Chinese satellite also launched by China.
- SARAL (2013-009A) is a satellite launched by India. The altimeter AltiKa and DORIS are French contributions.
- CryoSat-2 (2010-013A) is a satellite launched by the ESA to study polar sea ice. Of special interest is the *freeboard*, the amount by which the ice sticks out of the water. From this, the thickness, and, with the surface area, the total volume may be calculated. In-orbit positioning is done using the French DORIS system.

  The launch of CryoSat-1 failed.

  varalaita
- Sentinel-3A (2016-011A) is a versatile ESA remote-sensing satellite, the first of a planned constellation. It carries several instruments, among them the SRAL, or Synthetic Aperture Radar Altimeter.

  synteettinen aukko

The measurement method of satellite radar altimetry is presented in figure 13.2. The figure shows all the quantities playing a role in altimetry: the measured range s is the height h of the satellite above the reference ellipsoid corrected for the geoid height N, the sea-surface topography H, and variations of the sea surface, like tides, eddies, annual variation, and so on.

vertaus-ellipsoidi

Furthermore, if the satellite does not contain a precise positioning device, the true orbit of the satellite will differ from the calculated orbit

[2] Walter Heinrich Munk (1917–2019) was a famous American physical oceanographer.

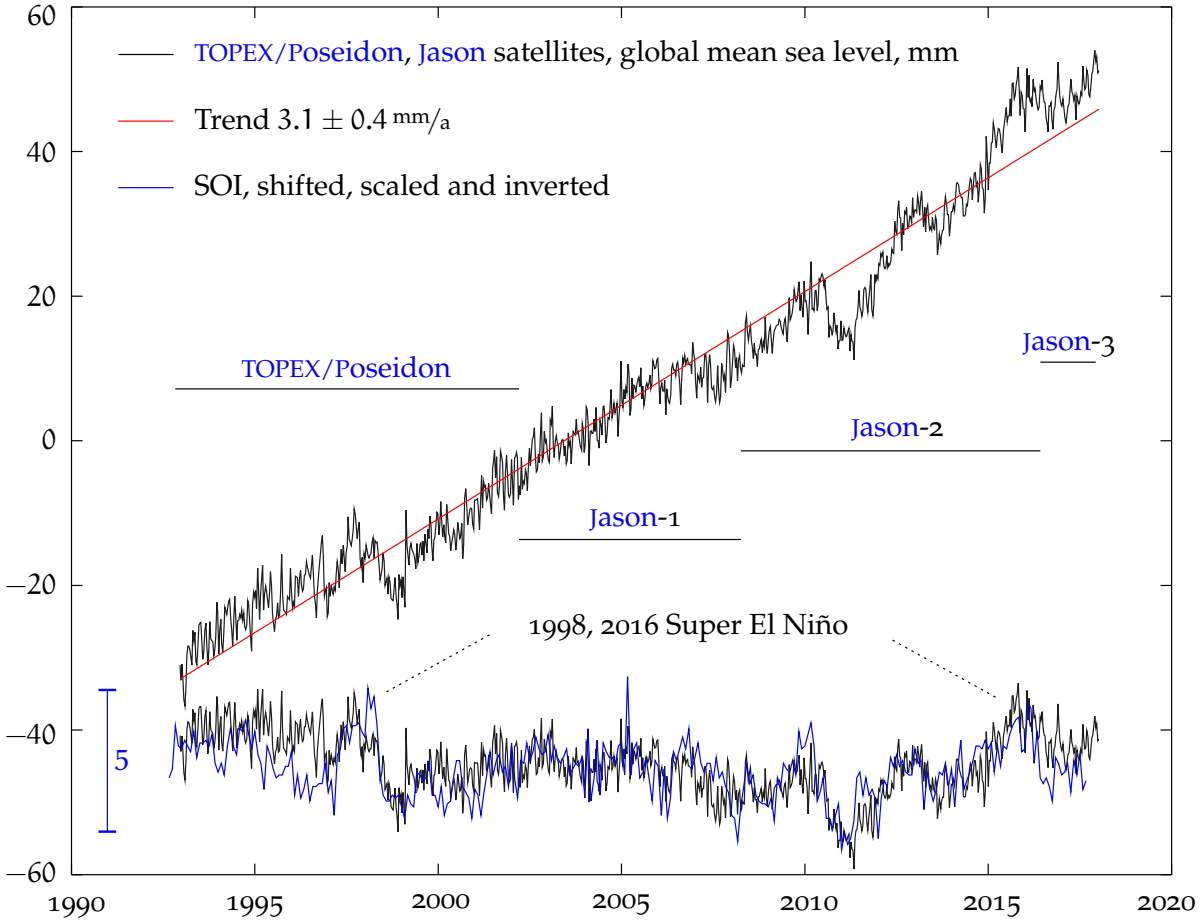


FIGURE 13.1. Results from the TOPEX/Poseidon and Jason satellites. Annual cycle removed. Data © Colorado University at Boulder's Sea Level Research Group; Nerem et al. (2010). Comparison with ENSO ("El Niño"), SOI = Southern Oscillation Index, Climate Research Unit, Climate Research Unit; Ropelewski and Jones (1987).

— even from the orbit calculated afterwards. Therefore,

$$h = h_0 + \Delta h,$$

in which h_0 is the calculated orbit, and Δh the orbit-error correction.

The measurements are performed by sending thousands of pulses down each second, measuring the travel times of the reflected return

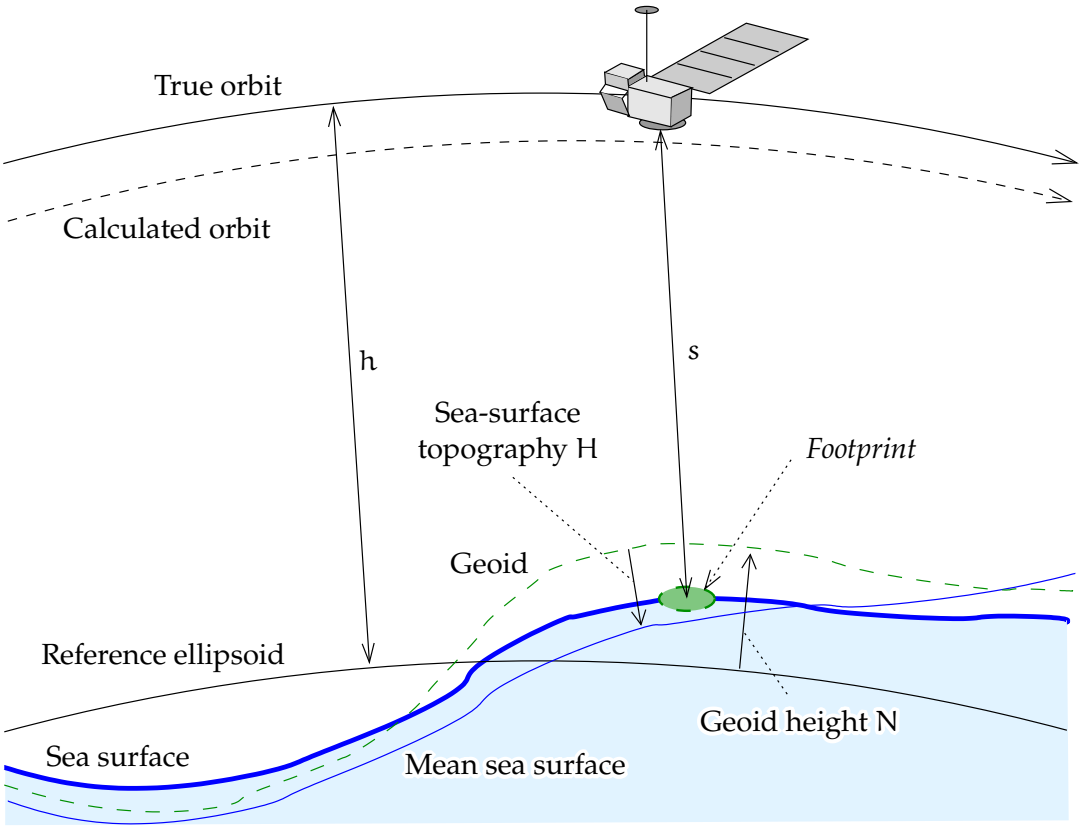


FIGURE 13.2. Satellite altimetry as a measurement method, concepts.

pulses on board the satellite, averaging them down to a measurement rate of 10–20 values per second, and transmitting these to Earth. Of these values, the largest and smallest are thrown away as possibly erroneous, and from the remainder, a mean value is calculated for the central epoch of the pulse train using linear regression. The value thus obtained from the regression line is the actual "measurement": one every second, making the effective measurement frequency 1 Hz.

The details will vary from satellite to satellite. The return pulse shape is never quite crisp, the place of the reflection on the ocean surface, or *footprint*, has a diameter of several kilometres. Especially if the ocean has wave motion (*significant wave height*, SWH), then, in the processing phase, one should make a careful correction so no bias is created: if

merkitsevä aallonkorkeus

the SWH is large, the altimeter footprint — the area on the sea from which radio energy returns to the receiver — will also be larger, and the distance travelled by the radio waves will on average be a little longer.

The newest satellites use an interferometric technique that differs somewhat from the description above.

Of all the corrections related to instrumentation, atmosphere, ocean, and solid Earth, we mention

1. the height of sea waves (SWH)
2. solid-Earth tides
3. ocean tides
4. the "wet" tropospheric propagation delay, best derived from measurements with a downlooking water vapour radiometer on the satellite, otherwise from meteorological modelling
5. the "dry" tropospheric propagation delay
6. the ionospheric delay, only for the part of the ionosphere below the satellite, depending on flight height
7. the altimeter's own calibration correction — nowadays "in-flight" calibration is always strived for, using an ensemble of GNSS-positioned tide gauges, see section 13.4.

mareografi

The measurements and all corrections to be made to them are collected into a "geophysical data record" (GDR), one per observation epoch. The files built this way are distributed to researchers. This allows all kind of experimentation; for example, the replacement of a correction by one calculated from improved models.

13.2 Crossover adjustment

When a satellite orbits the Earth over months or years, thousands of points are formed where the tracks cross each other. If we assume that the sea level is the same for both satellite overflights, then this forms a *condition* that can be used to adjust away orbit errors.

The observation equation is

$$\underline{s} = h - N - H - \epsilon + \underline{n} = h_0 + \Delta h - N - H - \epsilon + \underline{n},$$

in which $\underline{s}$ is the altimetric measurement of the height of the sea surface (including the known corrections 1–7 in the previous section), h the actual, and h_0 the calculated height of the satellite above the reference ellipsoid. N is the geoid height, H is the *sea-surface topography*: the permanent deviation of the sea surface from an equipotential surface, Δh is the orbit-error correction, ϵ is the *residual variation* of the sea surface, the variation remaining after correcting for the tides and other effects that can be modelled, and $\underline{n}$ is the random uncertainty, or noise, in the radar altimetry observations.

From this we obtain in the crossing point of tracks i and j:

$$\underline{\ell}_k \stackrel{\text{def}}{=} \left(\underline{s}^i - h_0^i\right) - \left(\underline{s}^j - h_0^j\right) = (\Delta h_i - \Delta h_j) - (\epsilon_i - \epsilon_j) + \left(\underline{n}_i - \underline{n}_j\right).$$

risteyskohta-tasoitus

This is the observation equation of crossover adjustment. Here we see the complication that sea-surface residual variation and orbit corrections appear in the equation in the same way. They cannot be separately determined by crossover adjustment.

If we forget for now the sea-surface residual variation — or assume that it behaves randomly, in other words it is part of the noise $\underline{n}$ — we may write more simply

$$\underline{\ell}_k = \Delta h_i - \Delta h_j + \underline{n}_k, \qquad \text{in which} \quad \underline{n}_k \stackrel{\text{def}}{=} \left(\underline{n}_i - \underline{n}_j\right) - (\epsilon_i - \epsilon_j).$$

The index k counts crossover points, the indices i and j count tracks.

Next, we choose a suitable *model* for the satellite orbit error. The simplest choice, sufficient for a small area, is the assumption that the orbit correction is a constant for each track. See a simple example, figure 13.3.

13.2.1 A simple example

In figure 13.3 we have three tracks and two crossing points. The *observation equations*, which describe the discrepancies in the known

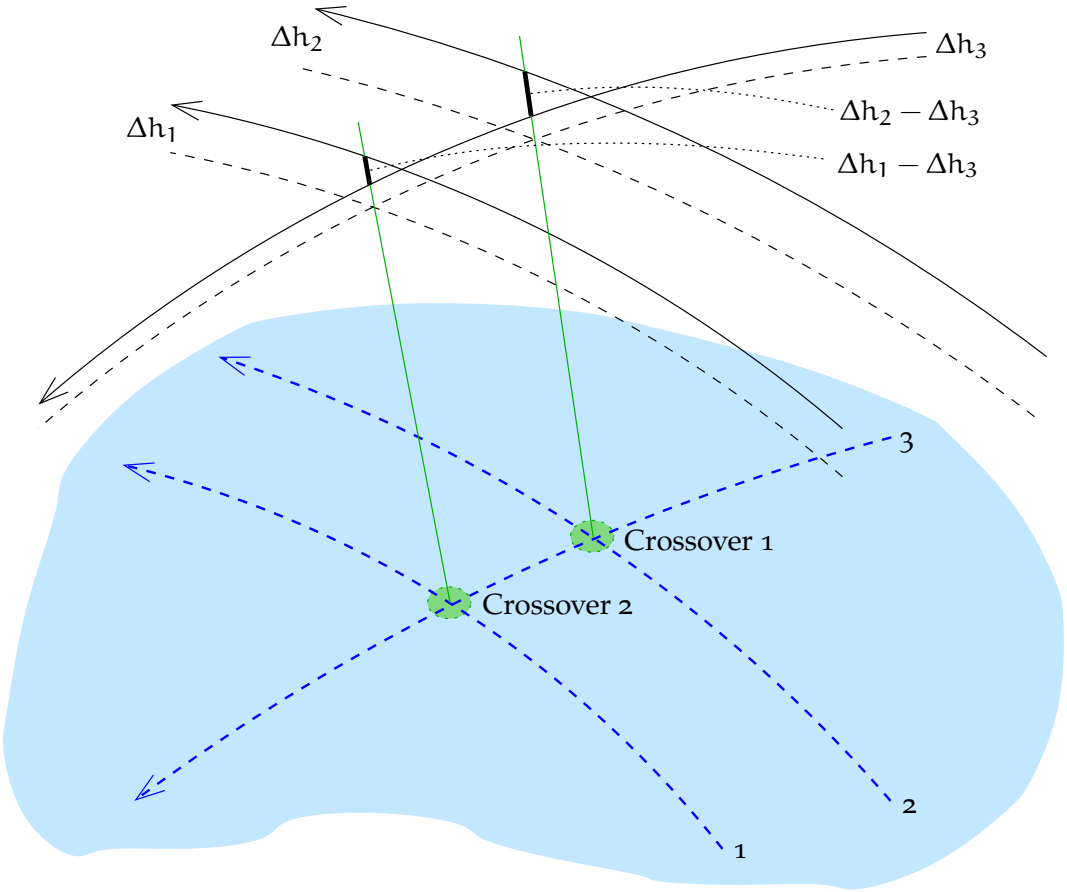


FIGURE 13.3. A simple crossover geometry.

crossover points as functions of the orbit corrections, are

$$\underline{\ell}_1 = \Delta h_2 - \Delta h_3 + \underline{n}_1,$$
$$\underline{\ell}_2 = \Delta h_1 - \Delta h_3 + \underline{n}_2,$$

3 or in matric form[3]

$$\overbrace{\begin{bmatrix} \underline{\ell}_1 \\ \underline{\ell}_2 \end{bmatrix}}^{\underline{\ell}} = \overbrace{\begin{bmatrix} 0 & 1 & -1 \\ 1 & 0 & -1 \end{bmatrix}}^{A} \overbrace{\begin{bmatrix} \Delta h_1 \\ \Delta h_2 \\ \Delta h_3 \end{bmatrix}}^{x} + \overbrace{\begin{bmatrix} \underline{n}_1 \\ \underline{n}_2 \end{bmatrix}}^{\underline{n}}, \tag{13.1}$$

symbolically

$$\underline{\ell} = A\mathbf{x} + \underline{\mathbf{n}}\,.$$

If one now tries to calculate the solution using ordinary least-squares,

$$\widehat{\mathbf{x}} = \left(A^{\mathsf{T}}A\right)^{-1} A^{\mathsf{T}}\underline{\ell}\,,$$

this will not work. The *normal matrix* $A^{\mathsf{T}}A$ is *singular* (check!). This makes sense, as one can move the whole track network up or down without the observations $\underline{\ell}_k$ changing. No unique solution can be found for such a system.

Finding a solution requires that *something* must be fixed: for example, one track — or, more democratically, the mean level of all tracks. This fixing is achieved by adding the following "observation equation":

$$\ell_3 \stackrel{\text{def}}{=} 0 = \begin{bmatrix} c & c & c \end{bmatrix} \cdot \mathbf{x}\,, \tag{13.2}$$

in which c is some suitable constant. Then, matrix A becomes

$$A = \begin{bmatrix} 0 & 1 & -1 \\ 1 & 0 & -1 \\ c & c & c \end{bmatrix},$$

and the least-squares solution

$$\widehat{\mathbf{x}} = \begin{bmatrix} \widehat{\Delta h}_1 \\ \widehat{\Delta h}_2 \\ \widehat{\Delta h}_3 \end{bmatrix} = \left(A^{\mathsf{T}}A\right)^{-1} A^{\mathsf{T}}\underline{\ell} = \left(A^{\mathsf{T}}A\right)^{-1} A^{\mathsf{T}} \begin{bmatrix} \underline{\ell}_1 \\ \underline{\ell}_2 \\ 0 \end{bmatrix},$$

where the matrix inversion is now possible. In this particular case, $\widehat{\mathbf{x}} = A^{-1}\underline{\ell}$ will give the same solution, as A is square and invertible:

$$\left(A^{\mathsf{T}}A\right)^{-1} A^{\mathsf{T}}\underline{\ell} = A^{-1}\left(A^{\mathsf{T}}\right)^{-1} A^{\mathsf{T}}\underline{\ell} = A^{-1}\left(\left(A^{\mathsf{T}}\right)^{-1} A^{\mathsf{T}}\right)\underline{\ell} = A^{-1}\underline{\ell}\,.$$

---

[3] Note the similarity with the observation equations for levelling! Instead of benchmarks, we have tracks, instead of levelling lines, crossover points.

Now the symbolic algebra system maxima (SourceForge, Maxima) — or brute-force calculation — gives the readily verified inverse

$$A^{-1} = \begin{bmatrix} 0 & 1 & -1 \\ 1 & 0 & -1 \\ c & c & c \end{bmatrix}^{-1} = \left(\begin{bmatrix} 1 & & \\ & 1 & \\ & & c \end{bmatrix} \begin{bmatrix} 0 & 1 & -1 \\ 1 & 0 & -1 \\ 1 & 1 & 1 \end{bmatrix} \right)^{-1} =$$

$$= \begin{bmatrix} 0 & 1 & -1 \\ 1 & 0 & -1 \\ 1 & 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 & & \\ & 1 & \\ & & c \end{bmatrix}^{-1} =$$

$$= \tfrac{1}{3} \begin{bmatrix} -1 & 2 & 1 \\ 2 & -1 & 1 \\ -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1/c \end{bmatrix} = \tfrac{1}{3} \begin{bmatrix} -1 & 2 & 1/c \\ 2 & -1 & 1/c \\ -1 & -1 & 1/c \end{bmatrix},$$

and the solution is

$$\begin{bmatrix} \widehat{\Delta h_1} \\ \widehat{\Delta h_2} \\ \widehat{\Delta h_3} \end{bmatrix} = A^{-1} \underline{\ell} =$$

$$= \tfrac{1}{3} \begin{bmatrix} -1 & 2 & 1/c \\ 2 & -1 & 1/c \\ -1 & -1 & 1/c \end{bmatrix} \begin{bmatrix} \underline{\ell}_1 \\ \underline{\ell}_2 \\ 0 \end{bmatrix} = \tfrac{1}{3} \begin{bmatrix} -1 & 2 \\ 2 & -1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} \underline{\ell}_1 \\ \underline{\ell}_2 \end{bmatrix},$$

from which c has vanished.

Another way to look at this is to first write the observation equations 13.1 and 13.2 together as

$$\overbrace{\begin{bmatrix} \underline{\ell}_1 \\ \underline{\ell}_2 \\ 0 \end{bmatrix}}^{\underline{\ell}} = \overbrace{\begin{bmatrix} 0 & 1 & -1 \\ 1 & 0 & -1 \\ c & c & c \end{bmatrix}}^{A} \overbrace{\begin{bmatrix} \Delta h_1 \\ \Delta h_2 \\ \Delta h_3 \end{bmatrix}}^{x} + \overbrace{\begin{bmatrix} \underline{n}_1 \\ \underline{n}_2 \\ 0 \end{bmatrix}}^{\underline{n}},$$

and then multiply the left-hand side and both terms on the right with the diagonal matrix

$$D \stackrel{\text{def}}{=} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/c \end{bmatrix}.$$

The result is

$$\overbrace{\begin{bmatrix} \underline{\ell}_1 \\ \underline{\ell}_2 \\ 0 \end{bmatrix}}^{D\underline{\ell}} = \overbrace{\begin{bmatrix} 0 & 1 & -1 \\ 1 & 0 & -1 \\ 1 & 1 & 1 \end{bmatrix}}^{DA} \begin{bmatrix} \Delta h_1 \\ \Delta h_2 \\ \Delta h_3 \end{bmatrix} + \overbrace{\begin{bmatrix} \underline{n}_1 \\ \underline{n}_2 \\ 0 \end{bmatrix}}^{D\underline{n}},$$

from which c has also vanished.

The principle applies generally:

> *Minimal constraints added to observation equations with a datum defect do not essentially change the solution.*

13.2.2 A more advanced orbit correction model

A more advanced representation of orbit corrections more suitable for use in a larger area, is a *linear function*:

$$\Delta h = a + b\tau,$$

where the parameter τ is the location along the track reckoned from its starting point. The dimension of this location can be time (seconds) or distance (degrees or kilometres). Now, the set of observation equations for the situation described above is

$$\overbrace{\begin{bmatrix} \underline{\ell}_1 \\ \underline{\ell}_2 \end{bmatrix}}^{\underline{\ell}} = \overbrace{\left[\begin{array}{cc|cc|cc} 0 & 0 & 1 & \tau_1^2 & -1 & -\tau_1^3 \\ 1 & \tau_2^1 & 0 & 0 & -1 & -\tau_2^3 \end{array}\right]}^{A} \overbrace{\left[\begin{array}{c} a_1 \\ b_1 \\ \hline a_2 \\ b_2 \\ \hline a_3 \\ b_3 \end{array}\right]}^{x} + \overbrace{\begin{bmatrix} \underline{n}_1 \\ \underline{n}_2 \end{bmatrix}}^{\underline{n}}.$$

The design matrix A contains, besides the values 1 and -1, also values $\pm\tau_k^i$, in which i is the number of the track and k that of the crossover point. These values are computable when the geometry of the tracks is known. rakennematriisi

Now there are *two* unknowns for every track, a and b, a constant and a trend. Of course also this system will prove to be singular. Removing the singularity can be done by fixing all three parameters b and one parameter a.[4]

[4] In order to understand this, build, say, a three-track "wire-frame model" from pieces of iron wire, tied together by pieces of string at the crossover points. The crossover conditions do not in any way fix the values of the trends b, and the whole absolute level of the frame continues to be unconstrained.

The phenomenon that no solution can be found unless something is fixed is called a *datum defect*. Fixing something suitable will define a certain *datum*. Between two different datums exists a *transformation formula*: in the case of one orbit correction parameter per track, this transformation is a simple parallel shift or *translation* of all tracks up or down.

korkeus-järjestelmä

The situation is somewhat similar to when one is defining a height or vertical reference system for a country. One needs to fix one height, for example that of Helsinki harbour. If alternatively one fixes another height, for example that of Turku harbour, the result is another *datum*, in which all height values differ from the corresponding ones in the first datum by a certain fixed amount.

The argument continues to hold if there is a large number of tracks: say, ten north-going and ten south-going tracks, crossing at 10×10 crossover points. Here, for two parameters per track, we would have 40 unknowns and no less than 100 observations. Still, we must constrain the absolute level and the various trends and possible other deformations of the whole network of tracks. A simple approach is to attach *a priori* uncertainties to the unknowns a_i and b_i to be derived, for example, from the known uncertainties of the orbit determination available. The least-squares adjustment equation then becomes

pienimmän neliösumman tasoitus

$$\hat{\mathbf{x}} = \left(A^T A + \sigma^2 \Sigma^{-1}\right)^{-1} A^T \underline{\boldsymbol{\ell}},$$

in which Σ is the diagonal matrix containing the *a priori* variances[5] $\sigma^2_{a,i}$

[5]

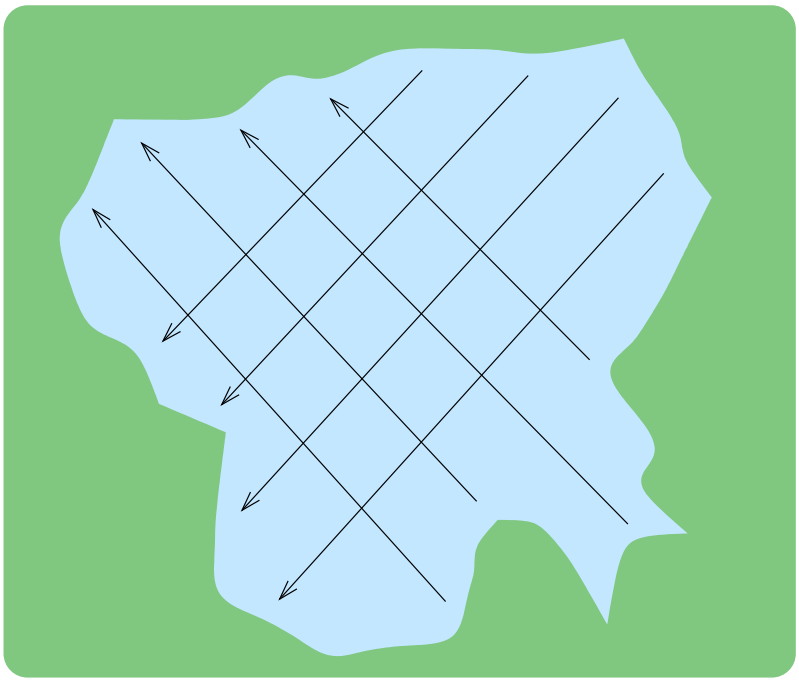


FIGURE 13.4. Example of track geometry of satellite altimetry.

and $\sigma^2_{b,i}$ of the parameters of each track i. This approach is referred to as Tikhonov[6] regularisation. 6

13.2.3 Another example

In figure 13.4 describing a satellite altimetry geometry, there are 16 crossover points. We attempt a crossover adjustment.

Questions

1. If the orbit correction Δh of each satellite track is described by a model with a single bias term, how many unknowns are there?

[5] σ is the mean error of unit weight, in this case the mean error, assumed constant, of a crossover observation.

[6] Andrey Nikolayevich Tikhonov (1906–1993) was a Russian mathematician and geophysicist.

2. If we have available 16 "observations" or crossover differences, how many of them are redundant?
3. Is it geometrically possible to calculate this network?
4. If we fix one track in advance (*a priori* information), how many redundant observations are there? Can this network be calculated?
5. If every track has *two* unknowns, a bias as well as a trend, a term changing linearly with time, what then needs to be fixed in order to make the network calculable? How many redundancies are there then?
6. If, in case 3, we fix one track, which one would you choose? Propose alternatively a solution where you do not have to make a choice.

Answers

1. As many as there are tracks: 8.
2. $16 - 8 = 8$.
3. No, because the absolute level of the whole network is indeterminate.
4. $16 - (8 - 1) = 9$. Now the network can be calculated.
5. If we assume that the tracks are straight in (x, y) co-ordinates, then the set of allowable transformations on the whole network is

$$\Delta h = a_{00} + a_{10}\, x + a_{01}\, y + a_{11}\, xy,$$

having four degrees of freedom. So, one needs to fix for example one bias and three trends, not all north- or all south-going. Then there are $16 - (16 - 4) = 4$ redundancies.
6. Any such choice would be arbitrary. Rather use the method described above, Tikhonov regularisation.

13.2.4 *Global crossover adjustment*

In a global crossover adjustment, often a still more sophisticated model is used,

$$\Delta h = a + b \sin\tau + c \cos\tau, \tag{13.3}$$

in which now τ is an angular measure, for example the place along the track reckoned from the last south-north equator crossing or *ascending node*. See Schrama (1989), where this problem is treated more extensively. In this model, a represents the size of the orbit, while b and c denote the offset of the centre of the orbit from the geocentre. This model is three-dimensional: the orbital arcs with their crossovers form a spherical network surrounding the Earth. The degrees of freedom left by the crossover conditions are now the size of this sphere and the offset of its centre from the geocentre:

$$\Delta h = a_0 + a_1 \cos\phi \cos\lambda + a_2 \cos\phi \sin\lambda + a_3 \sin\phi, \tag{13.4}$$

with four degrees of freedom.[7]

13.3 Choice of satellite orbit

In choosing a satellite orbit, Kepler's orbital laws are central. Kepler's third law says

$$GM_\oplus P^2 = 4\pi^2 a^3, \tag{13.5}$$

in which $a = a_\oplus + h$ is the satellite orbit's semimajor axis — the mean distance from the geocentre — while h is called the satellite's mean height. P is the orbital period; $a_\oplus$ is the equatorial radius of the Earth ellipsoid.

isoakselin puolikas

From equation 13.5 one can already infer that using satellite observations one can precisely determine the quantity $GM_\oplus$, the mass of the Earth multiplied by Newton's universal gravitational constant.[8] The

[7] One could argue that, in equation 13.3, the parameter a should be zero, as Kepler's third law allows a very precise determination of the orbital size, see section 13.3. Then, also $a_0 = 0$ in equation 13.4.

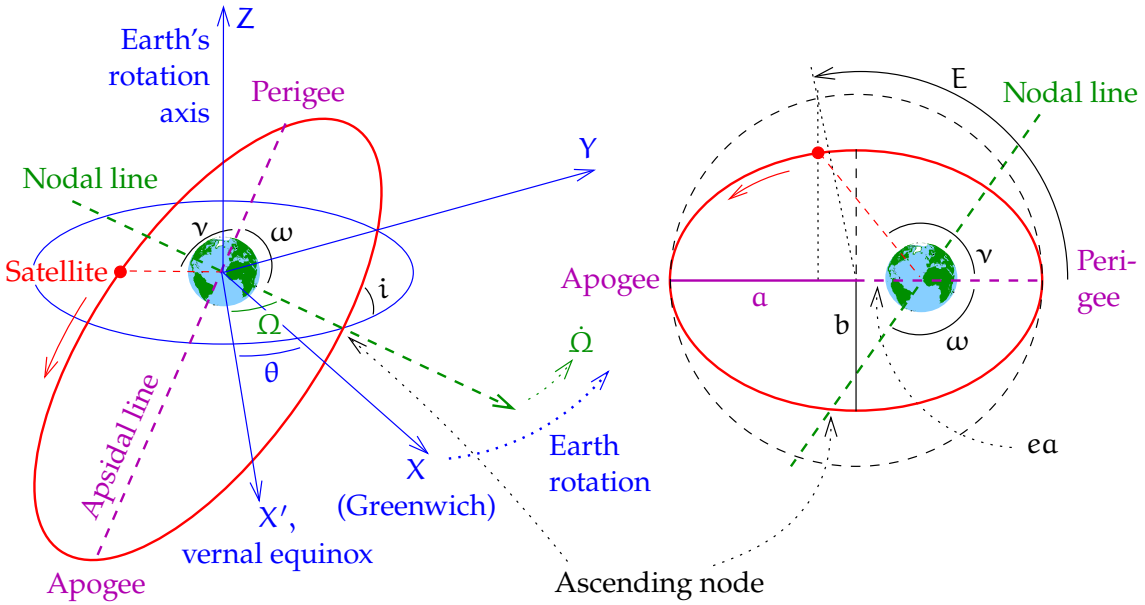


FIGURE 13.5. Kepler's orbital elements: a — semimajor axis, e — eccentricity, i — inclination, Ω — right ascension (celestial longitude) of the ascending node, ω — argument of perigee, and ν — true anomaly.

period P can be precisely determined from long observation series, and the size of the orbit a can also be obtained very precisely, for example from satellite laser ranging (SLR) observations. For this purpose the well-known LAGEOS (Laser Geodynamic Satellite) satellites (1976-039A and 1992-070B), which orbit the Earth at a height of 5900 km, have been used. Ranges are nowadays obtained with better than centimetre precision.

laseretäisyys-mittaus

The orbits of altimetric satellites are chosen to be much lower, as is seen from table 13.1 at the start of the chapter. The height is fine-tuned

[8] This is why it is said that Henry Cavendish was the first to "weigh the Earth"…. Determining $GM_\oplus$ was already straightforward back then using the orbital motion of the Moon, or even gravity on the Earth's surface. The challenge was separating G and the mass of the Earth $M_\oplus$ from each other, obtaining the latter in ordinary units of mass.

using on-board thrusters, so that the satellite passes over the same place, for example once a day, after 14 orbital periods. Alternatively one chooses an orbit that flies over the same place every third, seventeenth, 168^{th} day.... This is called the *repeat period*. rakettimoottori

The choice of the repeat period depends on the mission objective:

- If one wishes to study the precise shape of the *mean sea surface*, one chooses a *long* repeat period, in order to get the tracks as close together as possible on the Earth's surface.
- If one wishes to study the *variability* of the sea surface, one chooses an orbit that returns to the same location after a *short* time interval. Then, the grid of tracks on the Earth's surface will be sparser.

Parameters describing the figure of the Earth also affect satellite motion, for example the quantity J_2, the *dynamic flattening*, having a value of $J_2 = 1082.63 \cdot 10^{-6}$. It is the largest of the many spherical-harmonic coefficients that together represent the figure of the Earth and that affect satellite orbits. In the case of J_2, the effect is that the plane of the satellite orbit rotates at a certain angular rate around the Earth's rotation axis: *orbital* or *nodal precession*. This typically makes the satellite, if it flies over the same location the next day, do so several minutes earlier. For a circular orbit of radius a, the equation is pallofunktiokerroin

$$\frac{d\Omega}{dt} = -\tfrac{3}{2}\sqrt{\frac{GM_\oplus}{a^3}}\left(\frac{a_\oplus}{a}\right)^2 J_2 \cos i,$$

in which $a_\oplus$ is the equatorial radius of the Earth reference ellipsoid, $M_\oplus$ the mass of the Earth, and i the *inclination* of the orbital plane relative to the equator.

Substituting numerical values into this yields

$$\frac{d\Omega}{dt} = -1.318\,95 \cdot 10^{18}\,\mathrm{m^{3.5}\,s^{-1}} \cdot \frac{\cos i}{(a_\oplus + h)^{3.5}},$$

in which h is the *mean height* of the satellite orbit, conventionally above a sphere of size equatorial radius $a_\oplus$. If we substitute into this, say, the

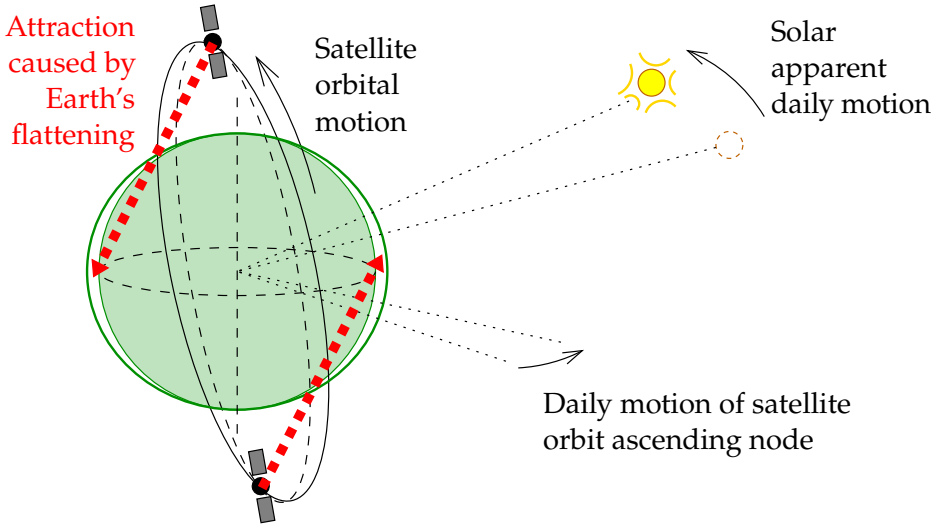


FIGURE 13.6. The mechanism of a Sun-stationary orbit.

satellite height $h = 800\,\mathrm{km}$ (and use $a_\oplus = 6\,378\,137\,\mathrm{m}$), we obtain

$$\frac{d\Omega}{dt} = -\,1.331\,03 \cdot 10^{-6}\,{}^{\mathrm{rad}}/_{\mathrm{s}} \cdot \cos i = \left(-6^\circ.589/_{\mathrm{day}}\right) \cdot \cos i.$$

For practical reasons — solar panels! — we often choose the satellite orbit such that the orbital plane turns along with the annual apparent motion of the Sun, ${}^{360^\circ}/_{365.25\,\mathrm{days}} = {}^{0^\circ.9856}/_{\mathrm{day}}$. See figure 13.6.

If the inclination i is chosen in the range 96°–102°, depending on the orbital height, then the Earth's dynamic flattening J_2 will cause just the suitable rotational motion of the orbital plane ("no-shadow / Sun-synchronous / Sun-stationary orbit"),[9] see figure 13.7.

An orbit with an inclination $i > 90^\circ$ is called a *retrograde orbit*: the satellite is moving westwards in longitude, opposite to the eastwards direction of the Earth's rotation. The orbital inclination i, or for a retrograde orbit, its supplement $180^\circ - i$, is also the greatest northern or southern geocentric latitude a satellite can fly over. This means that,

[9] If the height of the orbit is less than 1400 km, it cannot be completely no-shadow. In midwinter or in midsummer the satellite will then fly through the Earth's shadow.

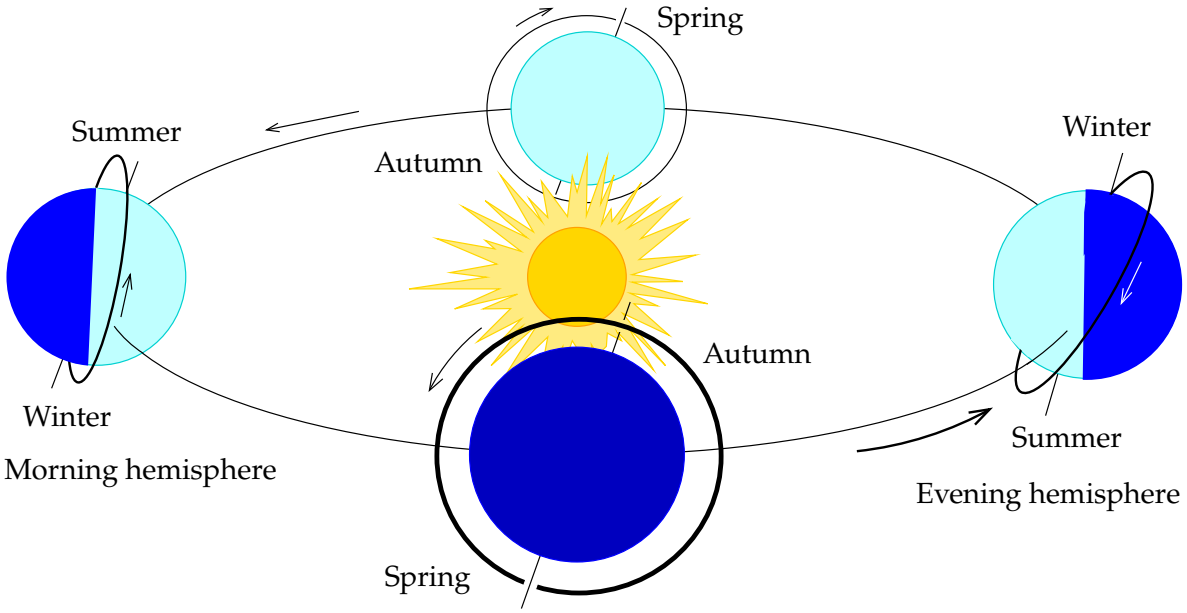


FIGURE 13.7. The geometry of a "no-shadow" orbit. In this figure, the satellite flies north over places where it is morning and south over places where it is evening.

unless the inclination is precisely 90°, there will be areas around both poles that the satellite will never overfly: the "polar holes".

A drawback of a Sun-stationary orbit is that the altimetric observations are always made at the same local time of day. For example, the diurnal and semidiurnal tides caused by the Sun will always have the same phase angle ("resonance"), and thus they cannot be observed with a satellite in this type of orbit. Therefore, the oceanographic satellite TOPEX/Poseidon, and the follow-up Jason satellites, were placed in non-Sun-stationary orbits.

13.3.1 Example

A satellite moves in a Sun-stationary orbit, in other words, always, day after day, flies over the same latitude at the same local mean solar time.

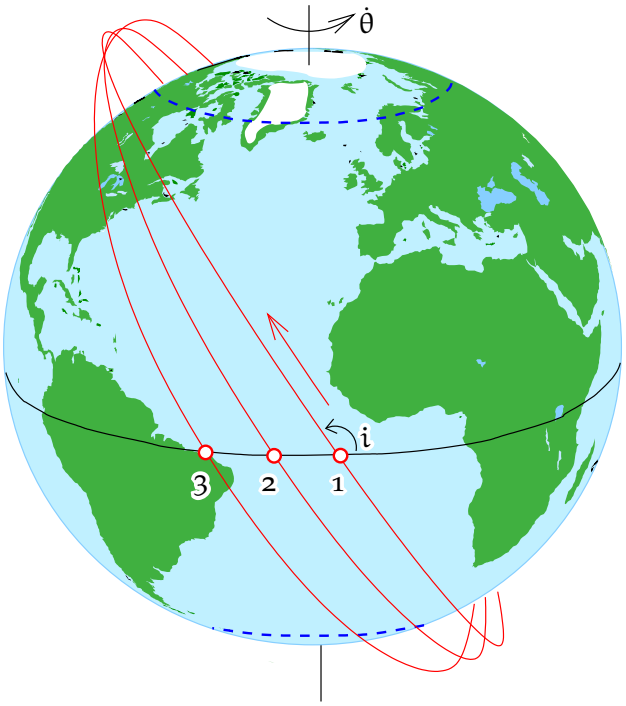


FIGURE 13.8. A satellite in a retrograde orbit around the rotating Earth, crossing the equator south to north three successive times. The angle between the orbit and the equator, the inclination i, or for a retrograde orbit, its supplement $180° - i$, is also the highest northern or southern latitude that the satellite can fly over. The unreachable "polar holes" are indicated by blue dashed lines.

Questions

1. What is the period of the satellite if it always flies again over the same spot after 14 revolutions?
2. The same question if the satellite always flies over the same spot after 43 revolutions (3 days)?
3. And after 502 revolutions (35 days)?
4. What is the height of the satellite in a "three-day orbit"? Use Kepler's third law, equation 13.5. $GM_\oplus = 3\,986\,005 \cdot 10^8\ \mathrm{m^3/s^2}$, and the height of the satellite is $h = a - a_\oplus$, with $a_\oplus =$

6 378 137 m.

5. What is the satellite height in a "35-day orbit"? And the height *difference* from the previous question?
6. What is, for the three-day orbit, the mean separation between north-going orbital tracks (so, at what level of detail is the altimeter able to image the sea surface!)?
7. The same question for a 35-day orbit.
8. Questions for reflection:
   (a) For what purpose would you use a 35-day orbit, for what purpose a three-day orbit?
   (b) Would it be possible, or easy, to fly both orbits with the same satellite (see question 5)?

Answers

1. The satellite completes 14 orbits per day, a day being 1440 minutes: $P = {}^{1440\,\text{min}}/_{14} = 102.857\,\text{min}$.
2. The satellite completes 43 orbits in three days or 3×1440 minutes: $P = {}^{3 \times 1440\,\text{min}}/_{43} = 100.465\,\text{min}$.
3. The satellite completes 502 orbits in 35 days or 35×1440 minutes: $P = {}^{35 \times 1440\,\text{min}}/_{502} = 100.398\,\text{min}$.
4. Execute the octave code in tableau 13.2. The result is 780.604 km.
5. The same code, with `P=100.398*60`, yields 777.421 km. The difference from the previous is 3.183 km.
6. There are 43 orbits with different ground tracks. That means a separation of ${}^{360^\circ}/_{43} = 8^\circ.372$. At the equator this is ${}^{40\,000\,\text{km}}/_{43} = 930\,\text{km}$. The distance is shorter at higher latitudes.
7. ${}^{360^\circ}/_{502} = 0^\circ.717$, or ${}^{40\,000\,\text{km}}/_{502} = 80\,\text{km}$.

TABLEAU 13.2. Calculating the height of a satellite from its period.

```
format long
GM=3986005e8;
ae=6378137;
P=100.465*60;
fac=4*pi*pi;
a=(GM*P*P/fac)^0.33333333;
h = a - ae;
printf('\n\nOrbital height: %8.3f km.\n', h/1000);
```

8.

(a) The 35-day orbit would be excellent for detailed mapping. The three-day orbit would be able to see, for example, tides or weather-related phenomena, albeit at poorer spatial resolution.

(b) The difference in height being only 3 km and in period 4 s, the change in orbit between the two repeat periods should be easily within reach of even small on-board thrusters.[10] So, yes.

10

13.4 In-flight calibration

The highly precise, GNSS-positioned satellite radar altimeters in use today require proper *calibration*. The technique of choice for this is *in-flight calibration*, using an ocean area — or sometimes a lake area — the geocentric location of the water surface of which is known thanks to surrounding GNSS-positioned tide gauges combined with a precise geoid model of the area. An example of such measurements is Vu et al. (2018).

[10] Disclosure: the total velocity change needed is $\Delta v = 1.6\,\mathrm{m/s}$, a brisk walking pace.

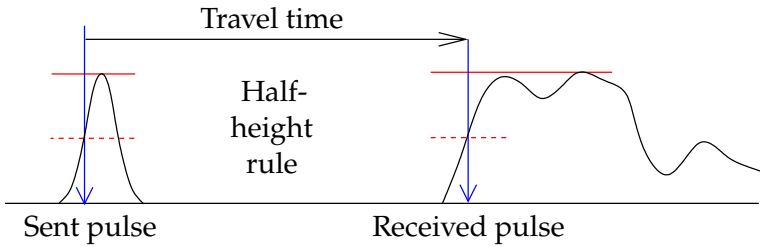


FIGURE 13.9. Analysing the altimeter return pulse. The classical return pulse time measurement uses the "half-height point".

One reason for in-flight calibration is the circumstance that radar altimeters not only have an unknown zero offset — due to the not precisely known signal paths through the electronic circuitry — but this offset may slowly change or drift over time, and may be temperature-dependent. käynti

13.5 Retracking

The results of a satellite altimetry mission are published already during flight in the form of a *geophysical data record* (GDR) file, containing everything related to the measurement, such as atmospheric correction terms, tidal corrections, and sea-state parameters.

It is common practice today to re-process older altimetry measurements, applying improved methodologies in order to extract additional useful information. The complete return pulse is analysed again in an approach called *retracking* (Altimetry, Retracking).

The method of analysis uses the point on the leading edge of the return pulse which is at half the height of the maximum value of the pulse. This is according to experience a good way to get the travel time associated with the point at the centre of the *footprint*, directly underneath the satellite. In the back part of the pulse are reflections from the further-away peripheral areas of the footprint.

There are three situations where the automatic analysis technique applied during flight does not work properly, and a more careful *a*

posteriori analysis of the pulse is worthwhile:

- Archipelagos like Indonesia or Åland. Here it may happen, for example, that the centre point of the footprint is on land. Then, the first strong bounces will come under an angle from the nearest coast. A precise coastline mask is then essential for processing. But already over open water close to coastlines the return pulse will be distorted.
- Sea ice areas in the Arctic and Antarctic Oceans. Bounces may come from the surface of the sea ice, in which case one should consider *freeboard* in the processing: how high the ice sticks out of the water.
- Over continental ice sheets. Here, the shape of the return pulse will be very different from that over open water. Furthermore, the travel time of the return pulse varies rapidly as the satellite flies on, and the reception window cannot keep track.[11]

mannerjäätikkö

In these cases the traditional real-time processing on-board produces erroneous measurements, or no measurements at all. With retracking, such measurements have been saved, and the area covered by altimetric measurements has been extended, especially into the Arctic and Antarctic areas.

Freeboard is an important quantity in determining the thickness of the ice. As the density of ice is about $920\,\mathrm{kg/m^3}$ and the density of sea water about $1030\,\mathrm{kg/m^3}$, the ice thickness is about $8\times$ freeboard.[12] If there is additional remote-sensing data providing the area of ice cover, one can calculate the total volume and mass of the sea ice.

The Arctic ice cover has diminished radically over recent decades. The most radical reduction has been that of ice volume, see figure

[11] The newest satellites such as Sentinel-3 use a digital terrain model for steering the tracking window when not over the open ocean.

[12] Assuming that there is no snow on the ice. Also, ice density varies, and differs between one-year and multi-year ice.

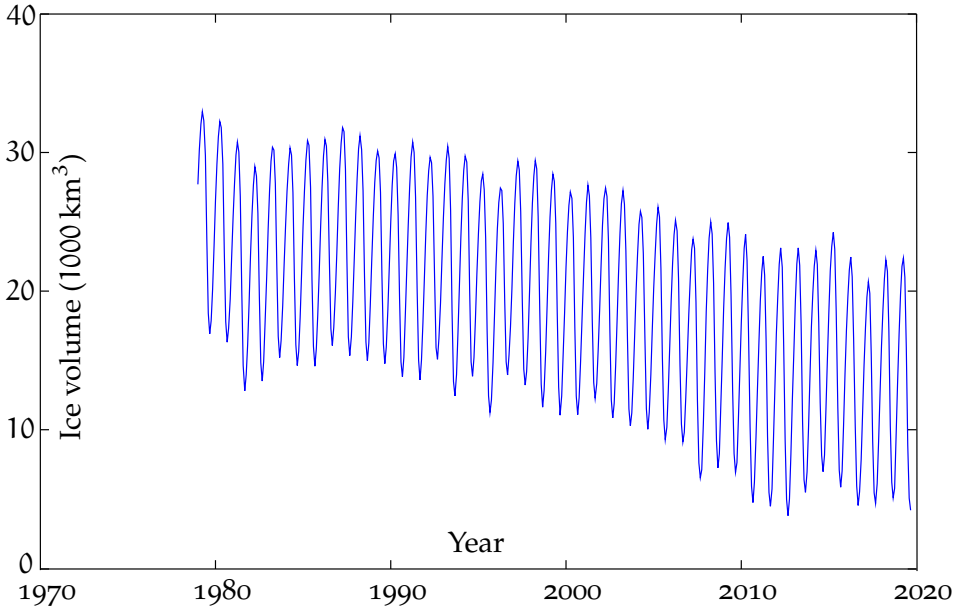


FIGURE 13.10. Ice volume on the Arctic Ocean. PIOMAS; Schweiger et al. (2011).

13.10. In addition to surface area, thickness is also decreasing: of the multi-year, thicker ice, a large part has already vanished.

13.6 Oceanographic research using satellite altimetry

The first geodetic application of satellite altimetry was geoid determination. Altimetric geoid determination works only if we assume that the sea surface

- is constant in time
- coincides with a level surface, the geoid.

In practice, however, the ocean surface is variable in time and is also not a level surface. Therefore, other approaches have been developed.

- Sea-surface variability can be studied by satellite altimetry using three methods:

- *Repeat tracks* from the same satellite. The tracks can be stacked and adjusted together using a simple orbit-error correction model, and the remaining per-track residuals tell something (though not everything) about the variability of the sea surface.
- *Crossovers* may also provide information on sea-surface variability. When the sea surface varies, the results from the crossover adjustment will get poorer: the root-mean-square of *a posteriori* (after calculation) crossover differences will grow. Using this method to actually *study* sea-surface variability is more challenging. It can however be used to estimate the magnitude of variability.
- Nowadays altimetric satellites always carry a GNSS *positioning instrument*, providing the absolute geocentric location of the microwave radar device at the moment of measurement. With it, the variations of sea level can be monitored by direct measurement, assuming that both temporal and spatial measurement densities are sufficient.

◦ The deviations of the sea level from a level surface — the geoid — can be studied only if we have access to independent information on the *true* geoid surface. If dense, high-quality gravity measurements are available for an area, these may be used to estimate the geoid, and after that one may calculate the sea-surface topography. Collecting sufficiently precise and dense gravimetric data is possible with sea or airborne gravimetry. Measurement with a special satellite (gravitational gradiometry, GOCE satellite) has also long been planned and has finally been realised, see subsection 13.7.3.

13.7 Satellite gravity missions

During the early years of the 21st century, three satellite missions were launched to investigate the fine structure of the Earth's gravity field

or geopotential; in other words, to determine a global high-resolution model of the geoid.

13.7.1 CHAMP

CHAMP (Challenging Minisatellite Payload for Geophysical Research and Applications, 2000-039B) was a German satellite project under the auspices of the German Research Centre for Geosciences GFZ. The satellite was launched into orbit from Plesetsk, Russia, in 2000. The orbit height was initially 454 km, coming down over the time of the mission to some 300 km due to atmospheric drag. The orbital inclination was 87°. On 19th September, 2010, the satellite returned into the atmosphere. Project description: CHAMP Mission.

CHAMP contained a GPS receiver in order to determine the satellite location in space $\mathbf{x}(t)$ over time t. From successive satellite locations one may calculate the geometric acceleration $\mathbf{a}(t)$ by differentiation:

$$\mathbf{a}(t) = \frac{d^2}{dt^2}\mathbf{x}(t).$$

The differentiation is done numerically as presented in the part on airborne gravimetry, equation 11.8.

The satellite also contained an accelerometer, which eliminated the satellite's accelerations caused by the atmosphere's aerodynamic forces, the deviations from free-fall motion. Then, only the accelerations caused by the Earth's gravitational field remain, from which a precise global geopotential or geoid model may be calculated using the techniques described earlier.

kiihtyvyys-mittari

A number of global geopotential models based on CHAMP data have been calculated and published.

13.7.2 GRACE

GRACE (Gravity Recovery And Climate Experiment Mission, 2002-012 A and B) measured *temporal changes* in the Earth's gravity field

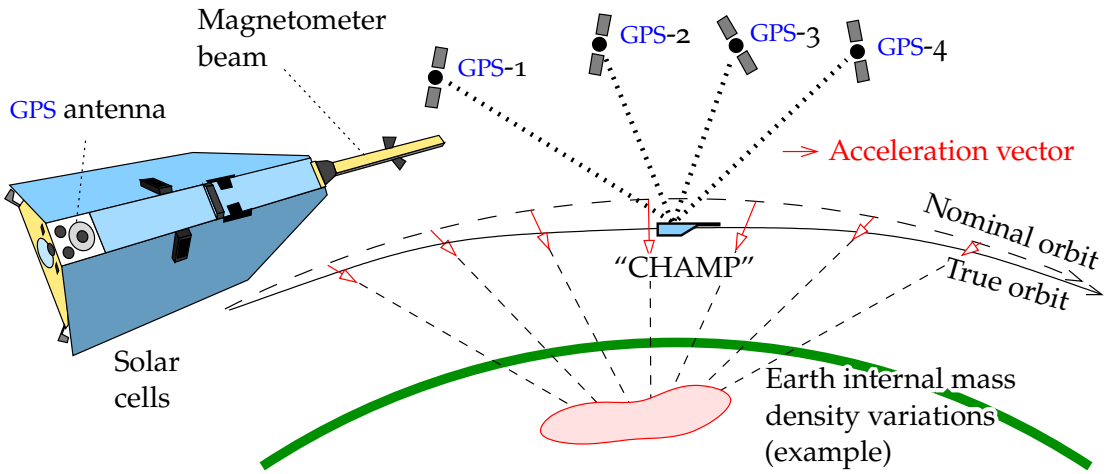


FIGURE 13.11. Determining the Earth's gravity field from GPS orbital tracking of a low flying satellite.

extremely precisely, but at a rather crude geographic resolution. These temporal changes are caused by motions in the Earth's "blue film": her atmosphere and hydrosphere. The quantity measured is also called the "sea-floor pressure", a somewhat surprising expression, until one sees that it really represents the total mass, variable in time, of a column of air and water.

The effective time resolution was one complete mapping of the globe every month. Project description: GRACE Mission. The project was a collaborative American-German undertaking under the leadership of the Center for Space Research, University of Texas at Austin.

GRACE was a *satellite pair* ("Tom and Jerry"): the satellites flew in the same orbit in a tandem configuration at initially about 500 km height, at an inter-satellite separation of 220 km. The orbital inclination was 89°, so the orbit was almost polar, providing complete global coverage. The changes in distance between the satellites were measured by a microwave link at a precision of $\pm 1\,\mu m/s$. Both satellites also carried sensitive accelerometers for measuring and eliminating the effect of atmospheric drag.

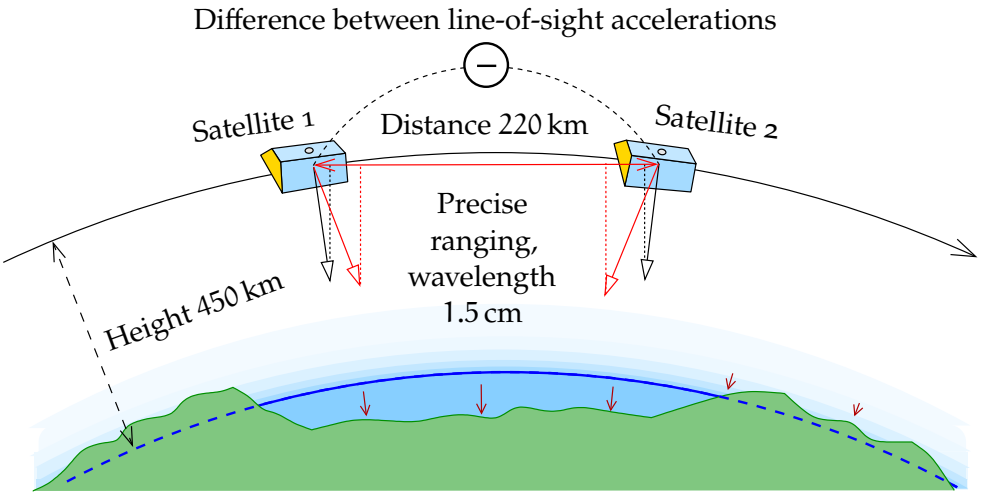


FIGURE 13.12. The principle of the GRACE satellites: measuring the minute variations in time of the gravity field using SST, satellite-to-satellite tracking. The changes are due to mass shifts in the "blue film" — the atmosphere and hydrosphere — and expressed as variations in "total sea-floor pressure" (↓).

The measurement system was so sensitive that even the movement of a water layer of one millimetre thickness could be noticed, as long as the layer extended over an area the size of a continent, some 1000 km.

The published results show impressively, for example, the wet and dry monsoons, seasonal variations in opposite phases in the northern and southern hemispheres, in the great tropical river basins: Amazonas, Congo, the Mekong, India, Indonesia. . . . GRACE Mission, hydrology.

The mission ended in 2017 after 15 years, three times the planned mission duration. A GRACE follow-on mission was launched in 2018, GRACE Follow-On Mission.

13.7.3 GOCE

GOCE (2009-013A, Geopotential and Steady-state Ocean Circulation Explorer) was the most ambitious of all the satellites. Built by the European Space Agency ESA, the satellite was launched successfully

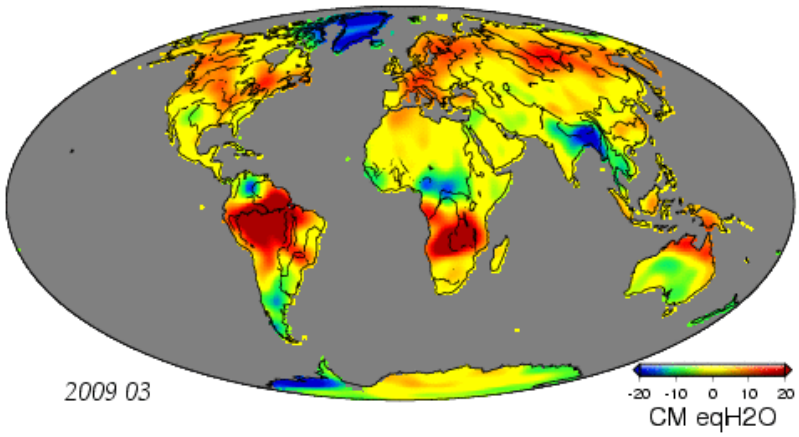


FIGURE 13.13. GRACE mission results: surface mass layer in centimetres of water equivalent. Click for animation (e-book).

from Plesetsk in March 2009. The orbital height was only 270–235 km during the mission and the satellite contained an ionic rocket engine with a stock of propellant in order to maintain the orbit against atmospheric drag. The orbital inclination was 96°.7, so the orbit was Sun-stationary.[13]

ajoaine

GOCE carried a very sensitive *gravitational gradiometer*, a device for measuring precisely components of the *gradient* of the Earth's attraction, the dependence of components of the attraction vector on the co-ordinates of place. The gradiometer consisted of six extremely sensitive, three-axes accelerometers mounted pairwise on a frame. The mission ended in 2013 and the satellite was seen to burn up in the atmosphere on 11th November over the Falkland Islands (Scuka, 2013).

Theoretical analysis has shown that a gravitational gradiometer is the best way to measure the very local features of the Earth's gravity field, better than orbital tracking by GNSS. The smallest details in the geoid map seen by GOCE are only some 100 km in diameter, and their

[13] Because of this inclination angle, there was a cap of radius 6°.7 at each pole within which no measurements were obtained. Over recent years these "poles of ignorance" have been gradually filled in by airborne gravimetry campaigns, for example Forsberg et al. (2017).

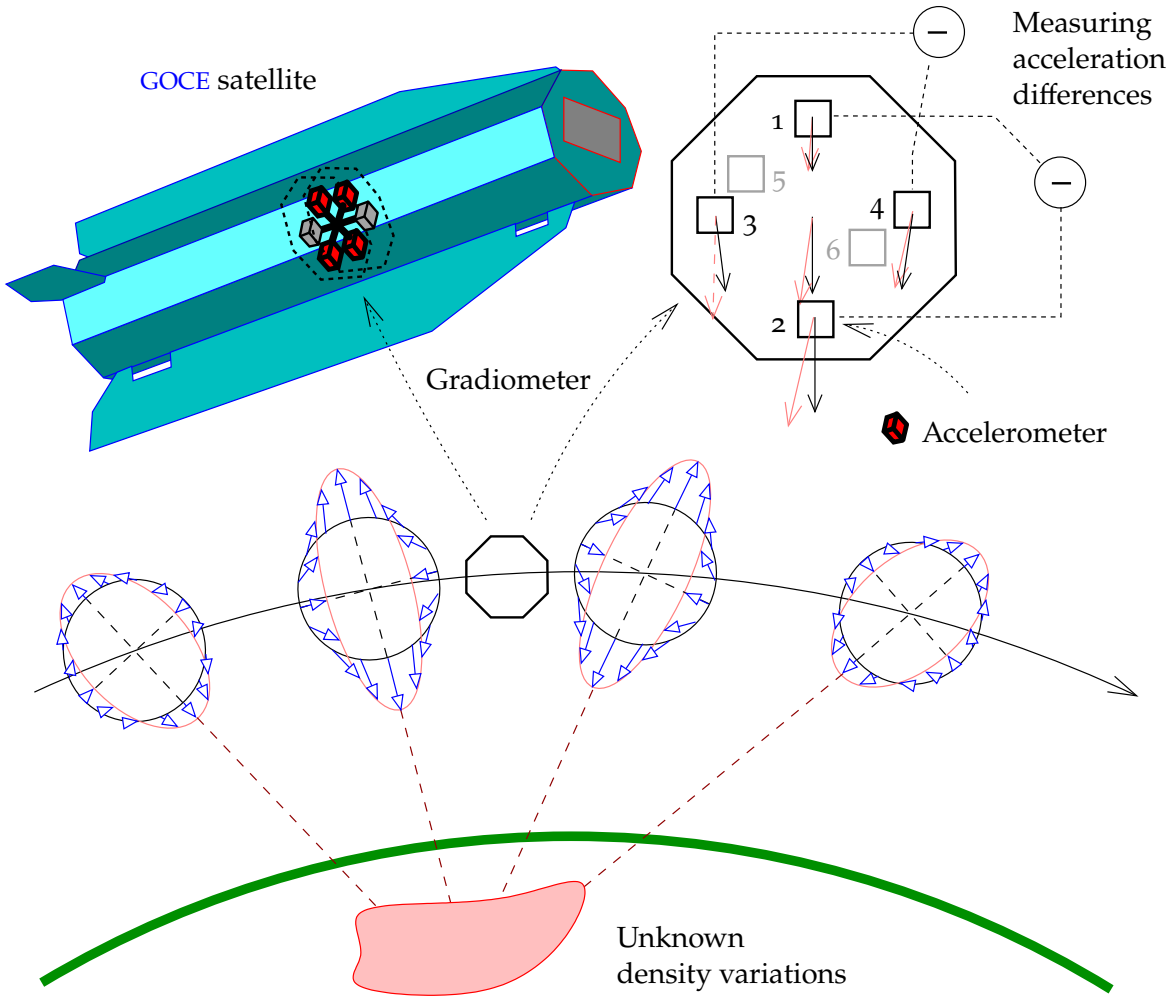


FIGURE 13.14. Determining the Earth's gravity field with the gravitational gradiometer on the GOCE satellite.

precision is as good as ± 2 cm.

With a global geoid model this precise, we may calculate the deviations of the sea surface from the geoid, an equipotential or level surface, at similar precision. We saw that the true location in space of the sea surface is obtained from satellite radar altimetry, also at a few centimetres' precision. This separation between sea surface and equipotential surface can again be inverted to *ocean currents*, see section

12.5 and figure 12.4. This is the background for the name of the GOCE satellite.

Self-test questions

1. What is the *footprint* of a radar altimeter? How does it depend on wave height?
2. What is the *freeboard* of ocean ice? How can it be used to determine the volume of the ice?
3. What three alternative models for the satellite orbit-error correction exist?
4. What is, in satellite altimetry crossover adjustment, a *datum defect*, and how can it be corrected?
5. How can Kepler's third law be used to determine the mean height of a satellite orbit if the satellite's period is given?
6. What is the repeat period of a satellite orbit?
7. What is J_2, and how does it affect the motion of a satellite?
8. What is a Sun-synchronous orbit, and why is it useful?
9. What is a retrograde orbit?
10. Why are the orbits of the TOPEX/Poseidon and Jason satellites not Sun-synchronous?
11. In table 13.1 some satellites have a repeat period that is an integer number of days, some satellites do not. What do satellites with non-integer repeat periods seem to have in common?
12. With which three satellite altimetric methods can one study sea-surface variability?
13. Three satellite missions have been launched so far to study the fine structure of the Earth's gravity field and its temporal variability. Present them and the methods used by them.

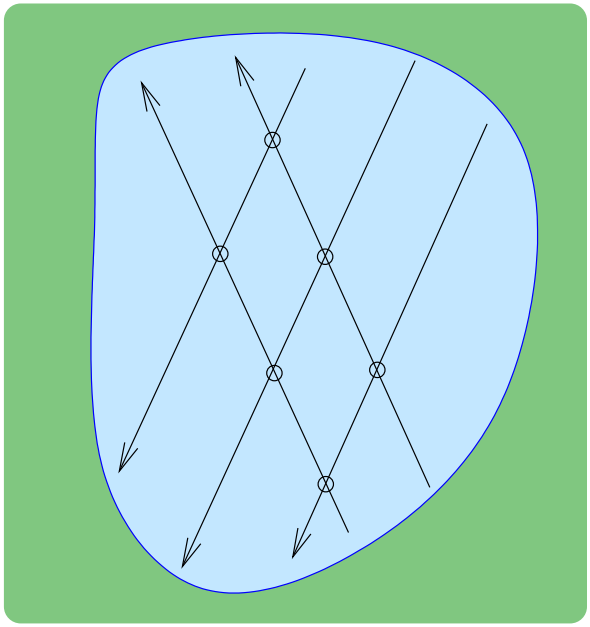


FIGURE 13.15. Example of a satellite altimetric track geometry.

Exercise 13–1: Altimetry, crossover adjustment

It is given that there are two north-going satellite tracks and three south-going ones. There are six crossovers, see figure 13.15.

1. If the orbit-error corrections for every track are described as a linear function of place:

$$\Delta h = a + b\tau,$$

   how many unknowns a and b are needed in total?

2. Write out the *observation equations*. The observations are the crossover differences, the unknowns are the coefficients a and b for the different tracks.

3. Can these observation equations give a unique solution? Why not?

4. How many unknowns need to be fixed for an unique solution? Which coefficients would you fix?

5. Are there any redundant observations? Was it wise to choose an orbit-error correction model with two unknowns per track?

Exercise 13–2: Satellite orbit

A satellite moves in a Sun-synchronous orbit. After 419 orbits and 30 days, the satellite again moves over exactly the same spot.

1. What is the period of the satellite?
2. How long is the distance (west to east), in kilometres, between the north-going tracks at the equator?
3. As a Sun-synchronous orbit is not polar, the highest northern latitude that the satellite can fly over is less than 90°. *In what compass direction* is the satellite flying at that point?

Exercise 13–3: Kepler's third law

1. What is the height h of a satellite of period 98 minutes? Use Kepler's third law 13.5,

$$GM_\oplus P^2 = 4\pi^2 a^3,$$

$GM_\oplus = 3\,986\,005 \cdot 10^8\,\mathrm{m^3/s^2}$, and the height of the satellite is $h = a - a_\oplus$, in which $a_\oplus = 6\,378\,137\,\mathrm{m}$.

2. What is the orbital inclination i of the satellite, if it is given that the orbit is circular and Sun-synchronous? See section 13.3.

Tides, the atmosphere, and Earth crustal movements

14

14.1 The theoretical tide

The tide is the result of the attraction of external celestial bodies, in the Earth's case Moon and Sun, and of the free-fall motion of the Earth towards these attracting bodies. We may write the potential field of the attraction as follows:

$$V' = \frac{GM}{\ell},$$

where ℓ is the distance of the attracting body from the point of evaluation of the potential, see figure 14.1. GM is the mass of the Sun or Moon multiplied by Newton's gravitational constant. The attraction can be expressed as an acceleration or "force" field, of magnitude

$$a' = \frac{GM}{\ell^2}.$$

The Earth, floating freely in space, responds to this by freely falling towards the attracting body, with an acceleration (see section 1.4):

$$a'' = \frac{GM}{d^2},$$

with d the distance of the attracting body from the Earth's centre.

This acceleration a'' is a constant. Considered as a field in space, we can associate a potential with it:

$$V'' = \frac{GM}{d^2} z = \frac{GM}{R} \left(\frac{R}{d}\right)^2 \frac{z}{R} = \frac{GM}{R} \left(\frac{R}{d}\right)^2 \cos\zeta,$$

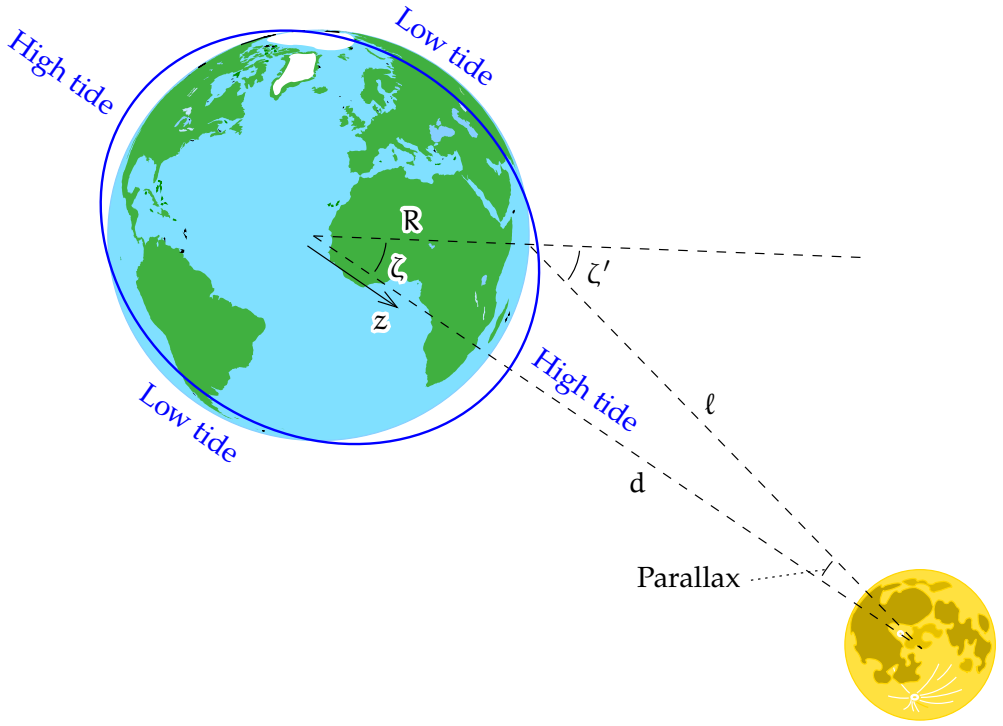


FIGURE 14.1. Theoretical tide. ζ' is the local zenith angle of the Moon (or Sun), ζ the corresponding geocentric angle.

where z is the co-ordinate defined along the line connecting the Earth's centre and the attracting body, see figure 14.1. R is the radius of the Earth, assumed spherical.

The net tidal potential as felt by mass elements on the Earth's surface is now

$$V = V' - V'' = \frac{GM}{\ell} - \frac{GM}{R}\left(\frac{R}{d}\right)^2 \cos\zeta = \\ = \frac{GM}{R}\sum_{n=0}^{\infty}\left(\frac{R}{d}\right)^{n+1} P_n(\cos\zeta) - \frac{GM}{R}\left(\frac{R}{d}\right)^2 \cos\zeta$$

using expansion 8.7. Here, the term for $n = 0$ is constant and thus arbitrary for a potential, and we delete it. The term $n = 1$ produces a

precise cancellation. Remains

$$V = \frac{GM}{R} \sum_{n=2}^{\infty} \left(\frac{R}{d}\right)^{n+1} P_n(\cos\zeta),$$

in which the term for degree 2 is dominant. asteluku

The tidal potential V can now be written as follows:

$$V = \frac{GMR^2}{d^3} P_2(\cos\zeta) + \cdots = \frac{GMR^2}{2d^3} \left(3\cos^2\zeta - 1\right) + \cdots,$$

ζ is the local *geocentric* zenith angle of the Sun or Moon: the local zenith angle ζ' corrected for parallax, see figure 14.1. $P_2(\cos\zeta)$ is the Legendre polynomial of degree two. In the case of the Sun and Moon, the extra terms $(\cdots)$ for higher degree numbers can be neglected, because these are such remote bodies: $d \gg R$.

The cosine rule on the sphere tells us that

$$\cos\zeta = \sin\phi \sin\delta + \cos\phi \cos\delta \cos h,$$

in which ϕ is the latitude, δ is the declination[1] of the Moon, and h is the hour angle[2] of the Moon. [1] [2]

The *spherical-harmonic addition theorem* (Wolfram MathWorld, Spherical Harmonic Addition Theorem) yields pallofunktioiden summauslause

$$P_n(\cos\zeta) = P_n(\sin\phi)\, P_n(\sin\delta) + \\ + 2 \sum_{m=1}^{n} \frac{(n-m)!}{(n+m)!} P_{nm}(\sin\phi)\, P_{nm}(\sin\delta) \cos mh,$$

or for $n = 2$,

$$P_2(\cos\zeta) = P_2(\sin\phi)\, P_2(\sin\delta) +$$

[1] The declination of a celestial body is its latitude on the celestial sphere, its angular distance from the celestial equator (Wikipedia, Declination), in this case as seen from the geocentre.

[2] The hour angle is the angle, or difference in longitude, between the meridian of the Moon and the local meridian, measured along the celestial equator (Wikipedia, Hour angle), in this case as seen from the geocentre. It vanishes when the Moon is in *upper culmination*, at its greatest elevation in the local meridian.

$$+\tfrac{1}{3}P_{21}(\sin\phi)\,P_{21}(\sin\delta)\cos h+\tfrac{1}{12}P_{22}(\sin\phi)\,P_{22}(\sin\delta)\cos 2h.$$

According to table 3.2,

$$\begin{aligned} P_{21}(\sin\phi) &= 3\sin\phi\cos\phi, & P_{21}(\sin\delta) &= 3\sin\delta\cos\delta,\\ P_{22}(\sin\phi) &= 3\cos^2\phi, & P_{22}(\sin\delta) &= 3\cos^2\delta, \end{aligned}$$

and we obtain

$$\begin{aligned} P_2(\cos\zeta) &= P_2(\sin\phi)\,P_2(\sin\delta)+\\ &\quad +3\sin\phi\cos\phi\sin\delta\cos\delta\cos h+\tfrac{3}{4}\cos^2\phi\cos^2\delta\cos 2h=\\ &=\tfrac{1}{2}\left(3\sin^2\phi-1\right)\tfrac{1}{2}\left(3\sin^2\delta-1\right)+\tfrac{3}{4}\sin 2\phi\sin 2\delta\cos h+\\ &\qquad +\tfrac{3}{4}\cos^2\phi\cos^2\delta\cos 2h. \end{aligned}$$

From this

$$V=\frac{GMR^2}{4d^3}\begin{pmatrix}\left(3\sin^2\phi-1\right)\left(3\sin^2\delta-1\right)+\\ +3\sin 2\phi\sin 2\delta\cos h+\\ +3\cos^2\phi\cos^2\delta\cos 2h\end{pmatrix}.$$

This is the *Laplace tidal decomposition equation*.

It has three parts:

- A slowly varying part,

$$V_1=\frac{GMR^2}{4d^3}\left(3\sin^2\phi-1\right)\left(3\sin^2\delta-1\right),$$

that still depends on the lunar declination δ and is therefore periodic with a 14-day (half-month) period. Using spherical trigonometry:

$$\begin{aligned} &\sin\delta=\sin\epsilon\sin\ell\\ &\qquad\Longrightarrow\quad \sin^2\delta=\sin^2\epsilon\sin^2\ell=\sin^2\epsilon\left(\tfrac{1}{2}-\tfrac{1}{2}\cos 2\ell\right), \end{aligned}\tag{14.1}$$

in which ℓ is the longitude of the Moon in its orbit, reckoned from the equator crossing, and ϵ is the inclination of the Moon's orbital

plane with respect to the equator, on average 23°.5 but varying between 18°.3 and 28°.6. Thus we obtain

$$V_1 = \frac{GMR^2}{4d^3}\left(3\sin^2\phi - 1\right)\left(3\sin^2\epsilon\left(\tfrac{1}{2} - \tfrac{1}{2}\cos 2\ell\right) - 1\right),$$

where we have used result 14.1. We split $V_1 = V_{1a} + V_{1b}$ into two parts, a constant[3] and a periodic, semi-monthly ("fortnightly") part:

$$V_{1a} = \frac{GMR^2}{4d^3}\left(3\sin^2\phi - 1\right)\left(\tfrac{3}{2}\sin^2\epsilon - 1\right), \qquad (14.2)$$
$$V_{1b} = -\frac{GMR^2}{4d^3}\left(3\sin^2\phi - 1\right)\left(\tfrac{3}{2}\sin^2\epsilon\,\cos 2\ell\right).$$

[3] In case of the Moon, not precisely, because $\epsilon_{☾}$ is (slowly) time-dependent.

- In addition, there are a couple of terms in which the Moon's hour angle h appears, periods roughly a day and roughly half a day:

$$V_2 = \frac{GMR^2}{4d^3}\cdot 3\sin 2\phi \sin 2\delta \cos h,$$
$$V_3 = \frac{GMR^2}{4d^3}\cdot 3\cos^2\phi\cos^2\delta\cos 2h.$$

In both, we have in addition to h, still δ as a "slow" variable. These equations could be written out as the sums of various functions of the longitude of the Moon ℓ.

Use basic trigonometry again, with equation 14.1:

$$\cos^2\delta = 1 - \sin^2\delta = 1 - \sin^2\epsilon\,\sin^2\ell = $$
$$= 1 - \sin^2\epsilon\left(\tfrac{1}{2} - \tfrac{1}{2}\cos 2\ell\right),$$
$$\cos 2\ell\cos 2h = \tfrac{1}{2}\left(\cos(2\ell + 2h) + \cos(2\ell - 2h)\right),$$
$$\sin 2\delta = 2\sin\delta\cos\delta = 2\sqrt{\sin^2\delta\left(1 - \sin^2\delta\right)} =$$
$$= 2\sin\epsilon\sqrt{\left(\tfrac{1}{2} - \tfrac{1}{2}\cos 2\ell\right)\left(1 - \sin^2\epsilon\left(\tfrac{1}{2} - \tfrac{1}{2}\cos 2\ell\right)\right)},$$

leading to a trigonometric expansion in lunar longitude ℓ, and so on. See for example Melchior's[4] famous book (1978).

[4]

TABLE 14.1. The various periods in the theoretical tide. The widely used symbols were standardised by George Darwin.

	Changing function	Period		Darwinsymbol		Name
		Moon	Sun	Moon	Sun	
V_{1a}	-	-	-	M_0	S_0	Permanent tide
V_{1b}	$\cos 2\ell$	14^d	182^d	*Mf*[a]	*Ssa*[b]	Declination tide
V_2	$\cos h$	24^h50^m	24^h	K_1, O_1	S_1, P_1	Diurnal
V_3	$\cos 2h$	12^h25^m	12^h	M_2	S_2	Semidiurnal

[a] Lunar fortnightly

[b] Solar semi-annual

The coefficient

$$D \stackrel{\text{def}}{=} \frac{3GMR^2}{4d^3} = \tfrac{3}{4}\frac{GM}{d}\left(\frac{R}{d}\right)^2, \tag{14.3}$$

[5] "Doodson's[5] constant, " is taken separately from the above equations. The value for the Moon equals $D_{☾} = 26.8\,\text{cm} \times \gamma$ and for the Sun $D_{\odot} = 12.3\,\text{cm} \times \gamma$, with $\gamma \approx 9.81\,\text{m/s}^2$. See figure 14.2.

[6] The *periods* are listed in table 14.1 with their Darwin[6] symbols.

In practice, the diurnal and semidiurnal tides can be divided further into many "spectral lines" close to each other, also because the lunar orbit, like the Earth's orbit, is significantly eccentric.

[4] Paul Jacques Léon Camille baron Melchior (1925–2004) was a Belgian geophysicist and Earth tides researcher and founder of the Walferdange underground laboratory for geodynamics in Luxembourg.

[5] Arthur Thomas Doodson FRS (1890–1968) was a British oceanographer, a pioneer of tidal theory, also involved in designing machines for computing the tides. He was completely deaf.

[6] Sir George Howard Darwin FRS FRSE (1845–1912) was an English astronomer and mathematician, son of Charles Darwin of *Origin of Species* fame.

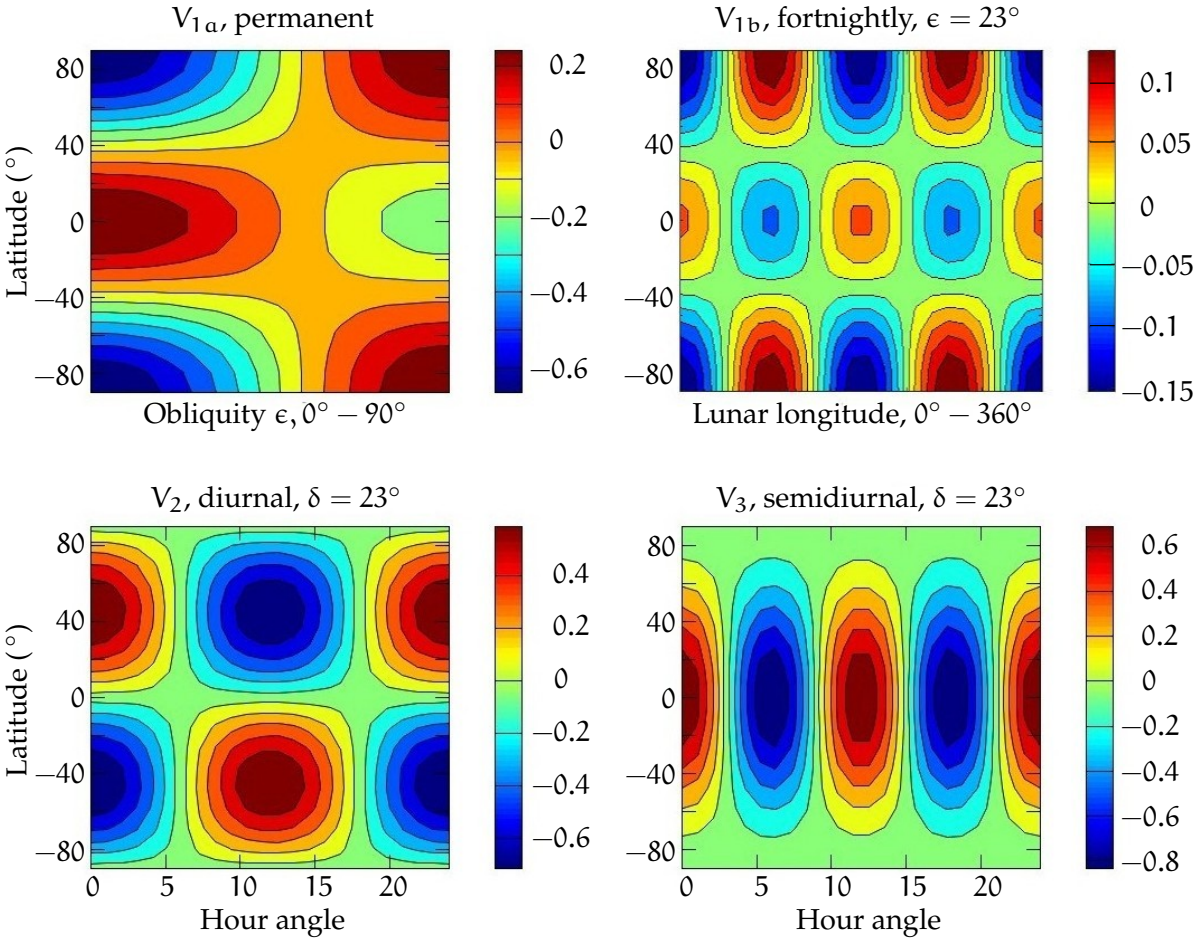


FIGURE 14.2. The main components of the theoretical tide. These values must still be multiplied by Doodson's constant D.

14.2 Deformation caused by the tidal potential

The tidal potential, or theoretical tide, of which we spoke above, is not the same as the deformation it causes in the solid Earth. This deformation will depend upon the elastic properties inside the Earth. These properties are often characterised by *elastic Love*[7] *numbers* (Love, [7]

[7] Augustus Edward Hough Love FRS (1863–1940) was a British mathematician and student of Earth elasticity.

1909; Melchior, 1978).

Let us first write the tidal potential $V = V(\phi, \lambda, r)$ in the following way:

$$V(\phi, \lambda, r) = \sum_{n=2}^{\infty} \left(\frac{r}{R}\right)^n V_n(\phi, \lambda) = \sum_{n=2}^{\infty} V_n^{\text{int}}(\phi, \lambda, r),$$

in which the index n denotes the spherical-harmonic degree number.

asteosuus $V_n(\phi, \lambda)$ is the degree constituent, and

$$V_n^{\text{int}}(\phi, \lambda, r) \stackrel{\text{def}}{=} \left(\frac{r}{R}\right)^n V_n(\phi, \lambda)$$

avaruus-pallofunktio the *interior* solid spherical harmonic of potential V for degree number n.

8 Call the linear[8] displacement of an element of matter of the solid Earth in the radial direction, u_r, in the north direction, u_ϕ, and in the east direction, u_λ. The following equations apply:

$$u_r(\phi, \lambda, r) = \frac{1}{\gamma} \sum_{n=2}^{\infty} H_n(r)\, V_n^{\text{int}}(\phi, \lambda, r) = \sum_{n=2}^{\infty} H_n(r)\, \zeta_n(\phi, \lambda, r),$$

$$u_\phi(\phi, \lambda, r) = \frac{1}{\gamma} \sum_{n=2}^{\infty} L_n(r) \frac{\partial V_n^{\text{int}}(\phi, \lambda, r)}{\partial \phi} = r \sum_{n=2}^{\infty} L_n(r)\, \xi_n(\phi, \lambda, r),$$

$$u_\lambda(\phi, \lambda, r) = \frac{1}{\gamma} \sum_{n=2}^{\infty} L_n(r) \frac{\partial V_n^{\text{int}}(\phi, \lambda, r)}{\cos\phi\, \partial \lambda} = r \sum_{n=2}^{\infty} L_n(r)\, \eta_n(\phi, \lambda, r).$$

Here, r is the distance from the geocentre. It is assumed here that the Love numbers H_n and L_n depend only on r, so that the elastic properties of the Earth are spherically symmetric. The symbols ζ_n, ξ_n and η_n represent the effect of the tidal potential of harmonic degree n on the level of an equipotential surface and on the components of the direction

luotiviiva of the plumb line.

epäsuora vaikutus The deformation of the Earth also causes a change, the "indirect effect"

[8] So their unit is metre, not degree, also for u_ϕ and u_λ!

in addition to the Moon's original tidal potential V, in the geopotential. We write

$$\delta V(\phi, \lambda, r) = \sum_{n=2}^{\infty} K_n(r)\, V_n^{\text{int}}(\phi, \lambda, r),$$

in which we already use a third type of Love numbers.

On the Earth's surface $r = R$ we make the following specialisation:

$$h_n \stackrel{\text{def}}{=} H_n(R), \qquad \ell_n \stackrel{\text{def}}{=} L_n(R), \qquad k_n \stackrel{\text{def}}{=} K_n(R). \tag{14.4}$$

Because of the large distances to Sun and Moon, the only non-negligible part of the tidal potential V is the part for the degree number $n = 2$, the "rugby-ball part" V_2^{int}.

The Love numbers will still depend on the *frequency*, or on the tidal *period* P:

$$h_n = h_n(P), \qquad \ell_n = \ell_n(P), \qquad k_n = k_n(P).$$

The tides offer an excellent means of determining all these Love numbers $h_2(P)$, $\ell_2(P)$, and $k_2(P)$ empirically, because, being periodic variations, they cause Earth deformations at the same periods, but with different amplitudes and phase angles.[9] In this way we may determine at least those Love numbers that correspond to periods occurring in the theoretical tide.

The numbers h and ℓ are nowadays obtained for example by GNSS positioning. The GNSS processing software contains a built-in reduction for this phenomenon. From gravity measurements one obtains information on a certain linear combination of h and k, $\delta = 1 + h - \frac{3}{2}k$: the lunar tidal force changes gravity directly, vertical displacement changes gravity though its gradient, and deformation of the Earth, the shifting of masses, also changes gravity directly.

The long water-tube tilt meter is also a useful research instrument, like the instrument of the Finnish Geodetic Institute that has long been

[9] The phase angles may be represented by making the Love numbers complex.

in use in the Tytyri limestone mine (Tytyri Mine Experience) in Lohja (Kääriäinen and Ruotsalainen, 1989). A modern, improved version of this instrument is presented in Ruotsalainen (2017). The same applies for sensitive clinometers in general, like the Verbaandert–Melchior pendulum. A clinometer measures the changes in orientation between the Earth's crust and the local plumb line. This can again provide information on a different linear combination of h and k, $\gamma = 1 - h + k$.

Measuring the *absolute* direction of the plumb line, for example with a zenith tube, can again provide information on the linear combination $\Lambda = 1 - \ell + k$, but only after various reductions (Earth orientation parameters like polar motion and variations in rotation rate), Vondrák et al. (2010). The Love number ℓ[10] comes in through the horizontal displacement of the zenith tube, to a location where the plumb-line direction is different.

[10]

14.3 The permanent part of the tide

As shown above, the theoretical tide equation contains a constant part that does not even vary in a long-period way. Of course the Earth also responds to this part of the tidal force. However, because the deformation is not periodic, it is not possible to measure it. And the mechanical theory of the elastic properties of the solid Earth, and our knowledge of the state of matter inside the Earth, are just not good enough for a theoretical calculation of the response.

For this reason the understanding is generally accepted that the effect of the permanent part of the tide on the Earth's state of deformation should not be included in any tidal reduction (Ekman, 1992). Many times, however, for example in the processing of GNSS observations or in defining spherical-harmonic expansions of the Earth's gravity field, the tidal reduction does include this term which it is theoretically and

pallofunktio-kehitelmä

[10] Also called the Shida number. Toshi Shida (1876–1936) was a Japanese Earth tide researcher.

practically impossible to know. See Poutanen et al. (1996).

More generally, the reduction of a geodetic quantity, for example the height of the geoid, for the permanent part of the tide can be carried out in three different ways:

- No reduction whatever is made for the permanent part. The quantity thus obtained is called the "*mean geoid*". The surface obtained is in the hydrodynamic sense an equilibrium surface, directly suitable for use in oceanography.
- The direct effect of the tidal field of Sun and Moon is removed in its entirety from the quantity, but the effect of the Earth's deformation caused by it is left uncorrected. The quantity thus obtained is called the "*zero geoid*".
- Both the effect of the tidal field of a celestial body, and the effect of the deformation it causes, can be calculated according to a certain deformation model (Love numbers), and removed. The result obtained is called the "*tide-free geoid*". Its problem is, as explained, the empirical indeterminacy of the elasticity model used.

See figure 14.3. It is good to be critical and precisely analyse the way in which the data reduction has been done!

14.4 Tidal corrections between height systems

We see from equation 14.2 that, with $\epsilon = 23^\circ.5$, the permanent part of the tidal potential is equal to

$$V_{\text{perm}} = \frac{GMR^2}{4d^3}\left(3\sin^2\phi - 1\right)\left(\tfrac{3}{2}\sin^2\epsilon - 1\right) \approx$$
$$\approx -0.7615 \cdot \frac{3GMR^2}{4d^3}\left(\sin^2\phi - \tfrac{1}{3}\right).$$

With the combined Doodson's constant 14.3 for Sun and Moon equal to

$$D = \frac{3GM_\odot R^2}{4d_\odot^3} + \frac{3GM_\leftmoon R^2}{4d_\leftmoon^3} =$$

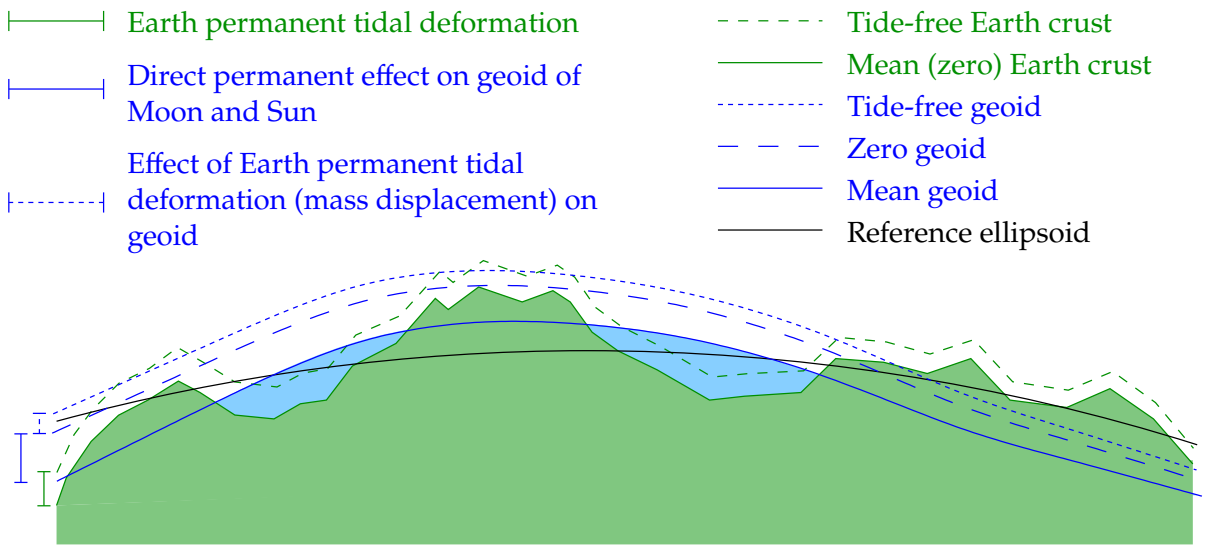


FIGURE 14.3. Conceptual diagram showing the constituents of the permanent tide.

$$= (12.3\,\text{cm} + 26.8\,\text{cm}) \times \gamma = 39.1\,\text{cm} \times \gamma$$

we obtain

$$V_{\text{perm}} = 29.77\,\text{cm} \times \left(\tfrac{1}{3} - \sin^2\phi\right) \times \gamma.$$

We can express this, with Bruns equation 5.2, as a permanent tidal geoid effect:

$$N_{\text{perm}} = 29.77\,\text{cm} \times \left(\tfrac{1}{3} - \sin^2\phi\right).$$

From this, $N_{\text{perm}}(0°) = 9.92\,\text{cm}$ on the equator, and $N_{\text{perm}}(\pm 90°) = -19.85\,\text{cm}$ on the poles.

This, the geoid effect of the permanent part of the external potential of the Sun and Moon, is also equal to the difference between the mean geoid and the zero geoid as defined above:

$$\Delta^{\text{mean}}_{\text{zero}} N \overset{\text{def}}{=} N_{\text{mean}} - N_{\text{zero}} = 29.77\,\text{cm} \times \left(\tfrac{1}{3} - \sin^2\phi\right).$$

For heights H above sea level, with $H = h - N$, we have

$$\Delta^{\text{mean}}_{\text{zero}} H \overset{\text{def}}{=} H_{\text{mean}} - H_{\text{zero}} = -29.77\,\text{cm} \times \left(\tfrac{1}{3} - \sin^2\phi\right),$$

and for two different latitudes ϕ_1 and ϕ_2 we have for the effect on the height difference

$$\Delta^{\text{mean}}_{\text{zero}} H(\phi_2) - \Delta^{\text{mean}}_{\text{zero}} H(\phi_1) = 29.77\,\text{cm} \times \left(\sin^2\phi_2 - \sin^2\phi_1\right).$$

This is the value to be *added* when going from a zero-geoid to a mean-geoid height system, and *subtracted* when going from a mean-geoid to a zero-geoid height system.

When the *tide-free* geoid and the Earth's crust enter into the picture, we need values for the Love numbers h and k for the permanent tidal deformation, expressing this deformation and its potential as fractional parts of the original external tidal potential.[11] As we have seen, these numbers cannot be empirically determined. Values often used are $h \approx 0.6$, $k \approx 0.3$. With this, the above equations apply with the coefficient 29.77 cm multiplied by the linear combination $\gamma = 1 - h + k \approx 0.7$. This yields

$$\Delta^{\text{mean}}_{\text{tidefree}} H \stackrel{\text{def}}{=} H_{\text{mean}} - H_{\text{tidefree}} = -20.84\,\text{cm} \times \left(\tfrac{1}{3} - \sin^2\phi\right),$$

$$\Delta^{\text{mean}}_{\text{tidefree}} H(\phi_2) - \Delta^{\text{mean}}_{\text{tidefree}} H(\phi_1) = 20.84\,\text{cm} \times \left(\sin^2\phi_2 - \sin^2\phi_1\right).$$

Any other correction equation can be obtained from these, like

$$\Delta^{\text{zero}}_{\text{tidefree}} H(\phi_2) - \Delta^{\text{zero}}_{\text{tidefree}} H(\phi_1) = -8.93\,\text{cm} \times \left(\sin^2\phi_2 - \sin^2\phi_1\right).$$

14.5 Loading of the Earth's crust by sea and atmosphere

In addition to the deformation caused by the tidal force, the Earth's crust also deforms due to the loading by sea and atmosphere. Especially close to the coast, the tidal motion of the sea causes a multi-period deformation that moves the Earth's crust up and down by as much as centimetres.

[11] Both are needed: the tidal deformation displaces both the Earth's crust and the geoid.

This phenomenon can be computationally modelled if the elastic properties of the solid Earth, the tidal motion of the sea, and the precise shape of the coastline are known. One known program for this purpose is the package Eterna written by the German Hans-Georg Wenzel,[12] which also has found use in Finland.

[12]

On the other hand, when such tools exist, tidal loading offers also an excellent opportunity for *studying* precisely the very local elastic properties of the Earth's crust.

For measuring the deformation, a registering gravimeter is generally used. The Earth's crust moves up and down elastically, which to first order changes gravity in proportion to the free-air gradient value $-0.3\,\mathrm{mGal/m}$. For a description of the method, see Torge (1992) section 4.2.

painovoiman ilmagradientti

The use of GNSS for measuring the ocean tidal loading has not yet become common.

Like the ocean, the atmosphere also causes, through changes in air pressure, varying deformations of the Earth's crust. The phenomenon is very small, at most a couple of centimetres. Gravity measurement is not a very good way to study this phenomenon, because many more local, often poorly known, factors affect local gravity. Measurement by GNSS is promising but also challenging.

Self-test questions

1. Present *in words* the three components of the theoretical tide produced by the Laplace decomposition method.
2. How may the slowly varying part of the theoretical tide be further decomposed into two parts? Present the parts *in words*.
3. What are the declination and hour angle of a celestial body, for example the Moon?

[12] Hans-Georg Wenzel (1945–1999) was a German physical geodesist and geophysicist.

4. What is Doodson's constant?
5. What do Love numbers express?
6. Why is it not possible to empirically determine the deformation caused by the permanent part of the tide?
7. Present the three different ways to take the permanent part of the tide into account when defining the geoid.

Exercise 14–1: The permanent tide

The equation for the permanent part of the tide is

$$V_{1a} = \frac{GMR^2}{4d^3}\left(3\sin^2\phi - 1\right)\left(\tfrac{3}{2}\sin^2\epsilon - 1\right),$$

in which ϕ is latitude and ϵ is the *obliquity* of the Earth's axis of rotation, currently about $23^\circ.5$.

1. For what value ϕ does the permanent part of the tide vanish? What is your interpretation?
2. For what value ϵ does the permanent part of the tide vanish? What is your interpretation?

Earth gravity field research

15

15.1 Internationally

In the framework of the IAG, the International Association of Geodesy, research into the Earth's gravity field is currently the responsibility of the International Gravity Field Service (IGFS). The IGFS was created in 2003 at the IUGG General Assembly in Sapporo, Japan, and it operates under the IAG's new Commission 2 "Gravity Field". The United States National Geospatial-Intelligence Agency (NGA) serves as its technical centre.

An important and well-reputed IAG service is the International Gravity Bureau, the BGI, *Bureau Gravimétrique International* located in Toulouse, France (http://bgi.obs-mip.fr/). The bureau works as an international broker to which countries can submit their gravimetric materials. If a researcher needs gravimetric material from another country, for example in order to do a geoid computation, they can request it from the BGI, who will provide it with the permission of the country of origin, provided the country of the researcher has in its turn submitted its own gravimetric materials for BGI use.

The French state has invested significant funds into this vital international activity.

Another important IAG service in this field is the ISG, the International Service for the Geoid. It has in fact been operating since as early as

1992 under the name International Geoid Service (IGeS), the executive arm of the International Geoid Commission (IGeC). The ISG office is located in Milan, Italy (http://www.isgeoid.polimi.it/). The task of this service is to support geoid determination in different countries. Existing geoid solutions are collected into a common database, and international research schools are organised to develop awareness about and skills in the art of geoid computation, especially in developing countries. The Italian state has provided significant funding for these activities.

Both services, BGI and ISG, are under the auspices of the IGFS, as two of the many official services of the IAG. Other IGFS services include the International Center for Earth Tides (ICET), the International Center for Global Earth Models (ICGEM), and the International Digital Elevation Model Service (IDEMS).

15.2 Europe

The EGU, the European Geosciences Union, operates in Europe, coordinating many publication and meeting activities relating to the gravity field and geoid. The EGU organises annual symposia, where sessions are always also included on subjects related to the gravity field and geoid. American scientists also participate. Conversely the
1 American Geophysical Union's (AGU) fall and spring meetings[1] are also favoured by European researchers.

The Geodetic Institute ("*Institut für Erdmessung*") of Leibniz University in Hannover, Germany has acted since 1990 as the computing centre of the International Geoid Commission's (IGeC) Subcommission for Europe, and produced high-quality European geoid models (Denker, 1998; European geoid calculations). The work continues since 2011 within the framework of the IAG Subcommission 2.4a *Gravity and Geoid in Europe*.

[1] Fall (autumn) meetings are in San Francisco, spring meetings somewhere in the world. The AGU, while American, is a very cosmopolitan player.

15.3 The Nordic countries

In the Nordic countries, important work is being co-ordinated by the NKG, the *Nordiska Kommissionen för Geodesi*, and its Working Group for Geoid and Height Systems. Its activities include geoid determination, studying the preconditions for still more precise geoid models, new levelling technologies, and the study of post-glacial land uplift.

The group has for a long time computed high-quality geoid models at its computing centre in Copenhagen, the next to last one being NKG2004 (Forsberg and Kaminskis, 1996; Forsberg and Strykowski, 2010). The newest model, NKG2015, is the result of calculations by the computing centres of several countries, including Sweden and Estonia. It was published in October 2016.

15.4 Finland

In Finland the study of the Earth's gravity field has mainly been in the hands of the Finnish Geodetic Institute, founded in 1918, one year after Finnish independence. The institute has been responsible for the national fundamental levelling and gravimetric networks and their international connections. In 2001 the Finnish Geodetic Institute's gravity and geodesy departments were joined into a new department of geodesy and geodynamics, to which gravity research also belongs.

Among topics studied are solid-Earth tides, the free oscillations of the solid Earth, post-glacial land uplift, and vertical reference or height systems.

korkeus-järjestelmä

Geoid models have been computed all the time, starting with Hirvonen's global model (Hirvonen, 1934) and ending, for now, with the Finnish model FIN2005N00 (Bilker-Koivula, 2010). These geoid models are actually based on the Nordic NKG2004 gravimetric geoid, and are fitted to a Finnish set of GNSS levelling control points, serving as a transformation surface for heights.

In 2015, the Finnish Geodetic Institute was merged into the National

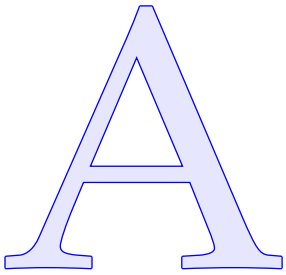
Land Survey as its geospatial data centre and research facility. The English-language acronym continues as FGI, the Finnish Geospatial Research Institute (https://www.maanmittauslaitos.fi/en/research).

Helsinki University of Technology (today part of Aalto University) has also been active in research on the Earth's gravity field. Heiskanen, a professor at HUT in 1928–1949, acted in 1936–1949 as the director of the International Isostatic Institute. After moving to Ohio State University, he worked with many other, including Finnish and Finnish-born, geodesists on calculating the first major global geoid model, the "Columbus geoid" (Kakkuri, 2008).

15.5 Textbooks

There are many good textbooks on the study of the Earth's gravity field. In addition to the already mentioned classic, Heiskanen and Moritz (1967), which is in large part obsolete, we may mention Wolfgang Torge's book (1989). Moritz (1980) is difficult but good. Similarly difficult is Molodensky et al. (1962). Worth reading also from the perspective of physical geodesy is Vaníček and Krakiwsky (1987). A newer book in the field is Hofmann-Wellenhof and Moritz (2006).

Field theory and vector calculus — core knowledge



A.1 Vector calculus

In physics, many quantities are *vector quantities*; for example, force, velocity, electrostatic field, and many more. The defining property of a vector is that under co-ordinate transformations it behaves identically to the location difference between two neighbouring points. Let the location difference be $\Delta\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$, in which $\mathbf{r}_1$ and $\mathbf{r}_2$ are the location vectors of points 1 and 2. In a co-ordinate transformation, the vector considered as an object does not change, but the numerical values of its *components*, subsection A.2.2, are co-ordinate system dependent and will change. The effect of the transformation on the components is the same as if the vector were a location difference between two points.

This is fundamentally why it is possible to *draw* vectors as *arrows*.

About notation In printed text, vectors are often written in bold: $\boldsymbol{v}$. In handwritten text one may use an arrow above the symbol: $\vec{v}$.

A.1.1 The scalar product

Between two vectors, a *scalar product* or dot product can be defined, which is itself a scalar value. A scalar is in physics a single numerical value; say, pressure or temperature. In the case of a scalar product of two vector fields, we speak of a *scalar field*: each scalar value is

tied to a location, but, even if a co-ordinate transformation changes the co-ordinate values of the location, the scalar value itself remains unchanged: it is an *invariant*.

An example of a scalar product: *work* ΔE is

$$\Delta E = \langle \mathbf{F} \cdot \Delta \mathbf{r} \rangle,$$

the scalar product of force $\mathbf{F}$ and path $\Delta \mathbf{r}$. Often, the angle brackets $\langle \cdot \rangle$ are left off.

Later we shall see that if the points 1 and 2, $\Delta \mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$, are very close to each other, we may write

$$dE = \langle \mathbf{F} \cdot d\mathbf{r} \rangle,$$

in which $d\mathbf{r}$ and dE are infinitesimal elements of path and energy. If now there is a curved path between points A and B, we may get from this an integral equation, the *work integral*:

$$\Delta E_{AB} = \int_A^B dE = \int_A^B \langle \mathbf{F} \cdot d\mathbf{r} \rangle.$$

A.1.2 The scalar product, formally

Let

$$s \stackrel{\text{def}}{=} \langle \mathbf{a} \cdot \mathbf{b} \rangle$$

be the scalar product of the vectors $\mathbf{a}$ and $\mathbf{b}$. It holds ($\mu \in \mathbb{R}$) that

$$\begin{aligned} \langle \mu\mathbf{a} \cdot \mathbf{b} \rangle = \langle \mathbf{a} \cdot \mu\mathbf{b} \rangle = \mu \langle \mathbf{a} \cdot \mathbf{b} \rangle, & \qquad \text{(homogeneity)} \\ \langle \mathbf{a} \cdot (\mathbf{b} + \mathbf{c}) \rangle = \langle \mathbf{a} \cdot \mathbf{b} \rangle + \langle \mathbf{a} \cdot \mathbf{c} \rangle, & \qquad \text{(distributivity)} \\ \langle \mathbf{a} \cdot \mathbf{b} \rangle = \langle \mathbf{b} \cdot \mathbf{a} \rangle, & \qquad \text{(commutativity)} \end{aligned}$$

and we call

$$\|\mathbf{a}\| \stackrel{\text{def}}{=} \sqrt{\langle \mathbf{a} \cdot \mathbf{a} \rangle}$$

the *norm* or length of vector $\mathbf{a}$.

The following also applies:

$$s = \|\mathbf{a}\| \, \|\mathbf{b}\| \cos \alpha,$$

where α is the angle between the directions of the vectors $\mathbf{a}$ and $\mathbf{b}$.

A.1.3 The exterior or vectorial product

The exterior product, or cross product, of two vectors is itself a vector called the *vectorial product*, at least in three-dimensional Euclidean space. For example, the *angular momentum* $\mathbf{L}$:

$$\mathbf{L} = \langle \mathbf{r} \times \mathbf{p} \rangle,$$

where $\mathbf{p} = m\dot{\mathbf{r}}$ is linear momentum, $\mathbf{r}$ the location vector of the body relative to some origin, m the mass of the body, and

$$\dot{\mathbf{r}} = \frac{d\mathbf{r}}{dt} \tag{A.1}$$

is the time derivative of the location, or *velocity*. We write

$$\mathbf{L} = m\langle \mathbf{r} \times \dot{\mathbf{r}} \rangle. \tag{A.2}$$

A.1.4 The vectorial product, formally

Let

$$\mathbf{x} \stackrel{\text{def}}{=} \langle \mathbf{a} \times \mathbf{b} \rangle$$

be the vectorial product of the two vectors $\mathbf{a}$ and $\mathbf{b}$. Then ($\mu \in \mathbb{R}$):

$$\begin{aligned}
\langle \mu\mathbf{a} \times \mathbf{b} \rangle = \langle \mathbf{a} \times \mu\mathbf{b} \rangle = \mu\langle \mathbf{a} \times \mathbf{b} \rangle, && \text{(homogeneity)} \\
\langle \mathbf{a} \times (\mathbf{b} + \mathbf{c}) \rangle = \langle \mathbf{a} \times \mathbf{b} \rangle + \langle \mathbf{a} \times \mathbf{c} \rangle, && \text{(distributivity)} \\
\langle \mathbf{a} \times \mathbf{b} \rangle = -\langle \mathbf{b} \times \mathbf{a} \rangle, && \text{(anticommutativity)}
\end{aligned}$$

and thus $\langle \mathbf{a} \times \mathbf{a} \rangle = 0$.

The resulting vector $\mathbf{x}$ is always perpendicular to the vectors $\mathbf{a}$ and $\mathbf{b}$. The length of vector $\mathbf{x}$ corresponds to the surface area of the parallelogram spanned by vectors $\mathbf{a}$ and $\mathbf{b}$: suunnikas

$$\|\mathbf{x}\| = \|\mathbf{a}\| \, \|\mathbf{b}\| \sin \alpha, \tag{A.3}$$

in which again α is the angle between the directions of vectors $\mathbf{a}$ and $\mathbf{b}$. If the angle is zero, then the vectorial product is also zero, because then, $\mathbf{a} = \mu\,\mathbf{b}$ for some suitable value of μ.

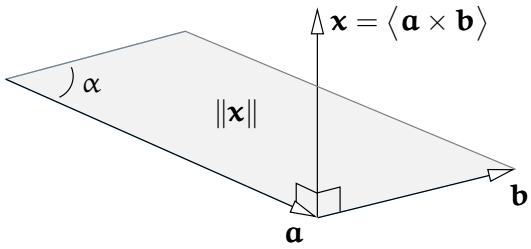


FIGURE A.1. Exterior or vectorial product.

If the angle is not zero we need in addition a *corkscrew rule* saying that, if a corkscrew is turned from vector $\mathbf{a}$ to vector $\mathbf{b}$, it will move forward in the direction of the product vector $\mathbf{x} = \langle \mathbf{a} \times \mathbf{b} \rangle$.

A.1.5 Kepler's second law

Let $\mathbf{r}$ be the location vector of the body (planet) relative to the centre of motion (the Sun), and $\dot{\mathbf{r}}$ (equation A.1) its velocity vector. Then, the vectorial product

$$\langle \mathbf{r} \times \dot{\mathbf{r}} \rangle = \left\langle \mathbf{r} \times \frac{d\mathbf{r}}{dt} \right\rangle \tag{A.4}$$

is precisely twice the surface area of the triangle or "area" swept over in a unit of time.

Let us take the time derivative of this product, the expression A.4:

$$\frac{d}{dt}\langle \mathbf{r} \times \dot{\mathbf{r}} \rangle = \left\langle \frac{d\mathbf{r}}{dt} \times \frac{d\mathbf{r}}{dt} \right\rangle + \left\langle \mathbf{r} \times \frac{d^2\mathbf{r}}{dt^2} \right\rangle = \langle \dot{\mathbf{r}} \times \dot{\mathbf{r}} \rangle + \langle \mathbf{r} \times \ddot{\mathbf{r}} \rangle. \tag{A.5}$$

Here, the first term vanishes, because for an arbitrary vector $\langle \mathbf{a} \times \mathbf{a} \rangle = 0$. In the second term, we can use our knowledge that the attractive force $\mathbf{F}$ emanating from the Sun that causes planetary orbital motion, and the associated acceleration,

$$\ddot{\mathbf{r}} = \frac{d^2\mathbf{r}}{dt^2},$$

keskeisvoima

are *central*:

$$\mathbf{F} = m\ddot{\mathbf{r}} = -\frac{GMm}{\|\mathbf{r}\|^3}\mathbf{r}.$$

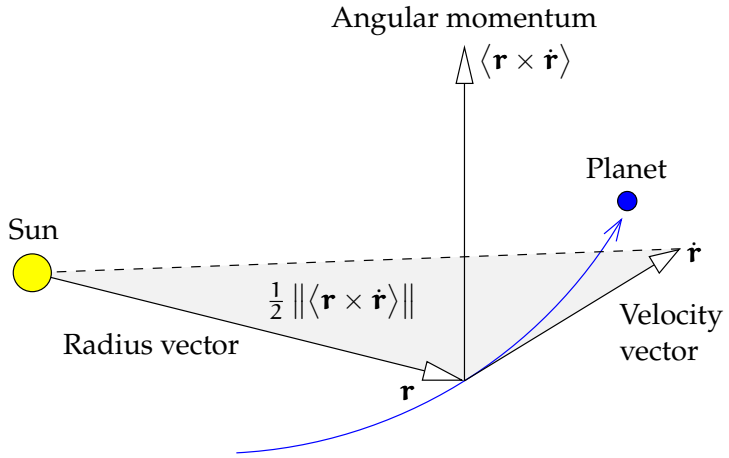


FIGURE A.2. Kepler's second law. In the same amount of time, the radius vector of a planet will "sweep over" a same-sized area — conservation of angular momentum.

G is the universal gravitational constant, M is the mass of the Sun, and m is the mass of the planet.

Substitute this into equation A.5:

$$\frac{d}{dt}\langle \mathbf{r} \times \dot{\mathbf{r}} \rangle = 0 - \frac{GM}{\|\mathbf{r}\|^3}\langle \mathbf{r} \times \mathbf{r} \rangle = 0.$$

So: the quantity on the left-hand side, angular momentum $\mathbf{L}$ per unit of mass m, equation A.2, is *conserved*:

$$\langle \mathbf{r} \times \dot{\mathbf{r}} \rangle = \frac{\mathbf{L}}{m}.$$

Like, for example, the total amount of energy, electric charge and many other quantities, the amount of angular momentum in a closed system is also constant.

A.2 Scalar and vector fields

A.2.1 Definitions

In the Euclidean space we may define functions or *fields*.

A scalar field is a scalar-valued function, which is defined throughout the space (or a part of it), for example temperature $T(\mathbf{r})$. So, for every value of the location vector $\mathbf{r}$ there is a temperature value $T(\mathbf{r})$.

A vector field is a vector-valued function that again is defined throughout space, for example the electrostatic field $\mathbf{E}(\mathbf{r})$.

A.2.2 A basis in space

In the space we may choose a *basis* made up of three vectors which *span* the space in question. Generally we choose three basis vectors $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$, that are orthogonal to each other, and the norms, or lengths, of which are equal to 1, an *orthonormal basis*. Orthogonality of two vectors means that their scalar product vanishes; so

ortonormaali kanta

$$\mathbf{i} \perp \mathbf{j}, \quad \mathbf{i} \perp \mathbf{k}, \quad \mathbf{j} \perp \mathbf{k}$$

means that

$$\langle \mathbf{i} \cdot \mathbf{j} \rangle = \langle \mathbf{i} \cdot \mathbf{k} \rangle = \langle \mathbf{j} \cdot \mathbf{k} \rangle = 0. \tag{A.6}$$

Orthonormality means in addition that

$$\|\mathbf{i}\| = \|\mathbf{j}\| = \|\mathbf{k}\| = 1. \tag{A.7}$$

Now we may expand vectors in the space into their *components*:

$$\mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k},$$

and scalar and vectorial products can now also be calculated with the aid of their components:

$$s = \langle \mathbf{a} \cdot \mathbf{b} \rangle = \langle (a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}) \cdot (b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}) \rangle =$$

$$= a_1b_1 + a_2b_2 + a_3b_3 = \sum_{i=1}^{3} a_ib_i,$$

using the identities stated above for the basis vectors A.6 and A.7.

For the vectorial product, the calculation is more involved. For orthogonal vectors, the angle α in equation A.3 is 90°, so

$$\|\langle \mathbf{i} \times \mathbf{j} \rangle\| = \|\langle \mathbf{i} \times \mathbf{k} \rangle\| = \|\langle \mathbf{j} \times \mathbf{k} \rangle\| = 1.$$

The corkscrew rule now tells us that

$$\begin{aligned} \mathbf{k} &= \langle \mathbf{i} \times \mathbf{j} \rangle = -\langle \mathbf{j} \times \mathbf{i} \rangle, \\ \mathbf{i} &= \langle \mathbf{j} \times \mathbf{k} \rangle = -\langle \mathbf{k} \times \mathbf{j} \rangle, \\ \mathbf{j} &= \langle \mathbf{k} \times \mathbf{i} \rangle = -\langle \mathbf{i} \times \mathbf{k} \rangle. \end{aligned}$$

We get as the final outcome the determinant

$$\begin{aligned} \mathbf{c} = \langle \mathbf{a} \times \mathbf{b} \rangle &= \det \begin{bmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{bmatrix} = \\ &= (a_2 b_3 - a_3 b_2)\,\mathbf{i} + (a_3 b_1 - a_1 b_3)\,\mathbf{j} + (a_1 b_2 - a_2 b_1)\,\mathbf{k}. \end{aligned}$$

So

$$c_1 = a_2 b_3 - a_3 b_2, \quad c_2 = a_3 b_1 - a_1 b_3, \quad c_3 = a_1 b_2 - a_2 b_1.$$

These expressions are determinants as well:

$$\begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} \det \begin{bmatrix} a_2 & a_3 \\ b_2 & b_3 \end{bmatrix} & \det \begin{bmatrix} a_3 & a_1 \\ b_3 & b_1 \end{bmatrix} & \det \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix} \end{bmatrix}^{\mathsf{T}}.$$

A.2.3 *The nabla operator*

The location vector $\mathbf{r}$ can be written on the $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ basis as follows: kanta

$$\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k},$$

which defines (x, y, z) co-ordinates in space.

Let us define a vector operator called *nabla* (∇) as follows:

$$\nabla \stackrel{\text{def}}{=} \mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z}.$$

The operator is on its own without meaning. It acquires meaning only when it *operates* on something, in which case the three partial derivatives on the right-hand side can be calculated.

A.2.4 The gradient

Let $V(\mathbf{r}) = V(x, y, z)$ be a scalar field in space. The nabla operator will give its *gradient* $\mathbf{g}$, a vector field in the same space:

$$\mathbf{g} = \operatorname{grad} V = \nabla V = \mathbf{i}\frac{\partial V}{\partial x} + \mathbf{j}\frac{\partial V}{\partial y} + \mathbf{k}\frac{\partial V}{\partial z}.$$

So, the field $\mathbf{g}(\mathbf{r}) = \mathbf{g}(x, y, z)$ is the gradient field of V. In physics, $\mathbf{g}$ is often a force field and V its potential.

Interpretation The gradient describes the *slope* of the scalar field. The direction of the vector is the direction in which the value of the scalar field changes fastest, and its length describes the rate of change with location. Imagine a hilly landscape: the height of the ground above sea level is the scalar field, and its gradient is pointing *uphill* everywhere, away from the valleys towards the hilltops. The longer the $\mathbf{g}$ arrows, the steeper the slope of the ground surface.

The gradient operator (like also the divergence and the curl, see later) is *linear*:

$$\operatorname{grad}(U + V) = \operatorname{grad} U + \operatorname{grad} V.$$

A.2.5 The divergence

Given is a vector field $\mathbf{a}(x, y, z) = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$. Form the *scalar product* s of this and the nabla operator:

$$s = \operatorname{div} \mathbf{a} = \langle \nabla \cdot \mathbf{a} \rangle = \frac{\partial a_1}{\partial x} + \frac{\partial a_2}{\partial y} + \frac{\partial a_3}{\partial z}.$$

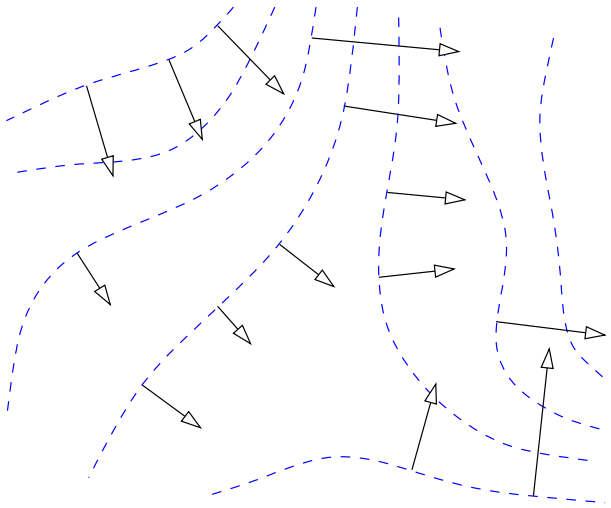


FIGURE A.3. The gradient. The level curves of the scalar field are dashed blue.

Interpretation The divergence describes the *sources* of a vector field, both the positive and negative ones. Think of the velocity of the flow of water as a vector field. At the locations of the "sources" the divergence is positive, at the locations of the "sewer holes" or *sinks*, negative; everywhere else zero (because liquid cannot appear out of nothing or disappear into nothing). lähteet, nielut

A.2.6 The curl

Given is again a vector field $\mathbf{a}(x, y, z)$. Form the *vectorial product* $\mathbf{c}$ of this and the nabla operator, again producing a vector field:

$$\mathbf{c} = \operatorname{curl} \mathbf{a} = \langle \nabla \times \mathbf{a} \rangle = \det \begin{bmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ a_1 & a_2 & a_3 \end{bmatrix} =$$

$$= \det \begin{bmatrix} \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ a_2 & a_3 \end{bmatrix} \mathbf{i} - \det \begin{bmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial z} \\ a_1 & a_3 \end{bmatrix} \mathbf{j} + \det \begin{bmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial y} \\ a_1 & a_2 \end{bmatrix} \mathbf{k} =$$

$$= \left(\frac{\partial a_3}{\partial y} - \frac{\partial a_2}{\partial z} \right) \mathbf{i} + \left(\frac{\partial a_1}{\partial z} - \frac{\partial a_3}{\partial x} \right) \mathbf{j} + \left(\frac{\partial a_2}{\partial x} - \frac{\partial a_1}{\partial y} \right) \mathbf{k},$$

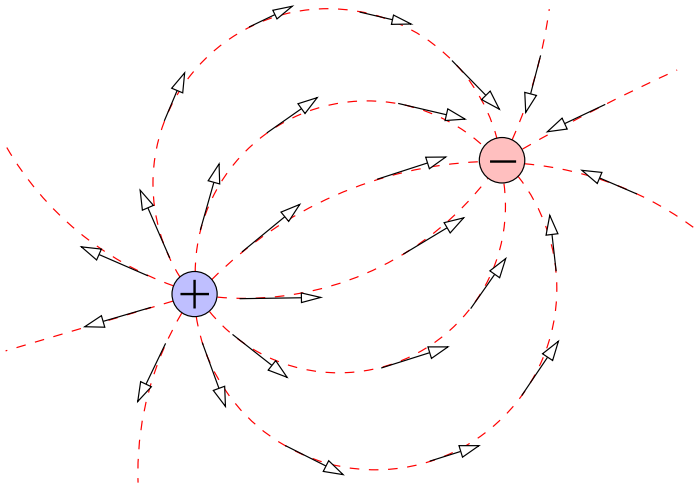


FIGURE A.4. The divergence. Positive divergences ("sources") and negative ones ("sinks"). Field lines red dashed.

using the evaluation rules for determinants.

Interpretation The curl describes the *eddiness* or *vorticity* or *turbulence* present in a vector field.

Imagine a weather map, where low- and high-pressure zones are drawn. Our vector field is the wind field. The wind circulates (in the northern hemisphere) clockwise around the high-pressure zones, and anticlockwise around the low-pressure zones. We may say that the curl of the wind field is positive at the high pressures and negative at the low pressures.

(This is a poor metaphor, as it is two-dimensional. In $\mathbb{R}^2$, the curl is a scalar, not a vector, just like we need only one angle to characterise a rotation, whereas in $\mathbb{R}^3$ we need the three Euler angles.)

A.2.7 Conservative fields

What happens if a vector field $\mathbf{a}$ is the gradient of a scalar field V, and

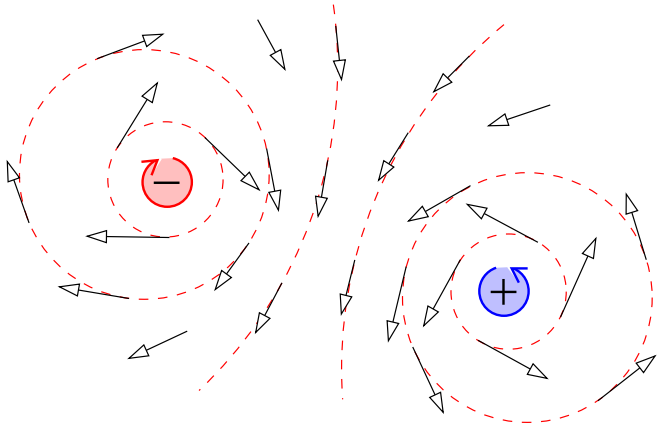


FIGURE A.5. The curl. Positive (anticlockwise) and negative (clockwise) eddies.

we try to calculate its curl $\mathbf{b}$, which is a vector as well? Write

$$\mathbf{b} = \operatorname{curl} \mathbf{a} = \operatorname{curl} \operatorname{grad} V = \det \begin{bmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \end{bmatrix} V$$

and let

$$\mathbf{b} = b_1 \mathbf{i} + b_2 \mathbf{j} + b_3 \mathbf{k}.$$

Then, expanding the determinant yields

$$\begin{aligned} b_1 &= \frac{\partial}{\partial y}\frac{\partial}{\partial z} V - \frac{\partial}{\partial z}\frac{\partial}{\partial y} V = 0, \\ b_2 &= \frac{\partial}{\partial z}\frac{\partial}{\partial x} V - \frac{\partial}{\partial x}\frac{\partial}{\partial z} V = 0, \\ b_3 &= \frac{\partial}{\partial x}\frac{\partial}{\partial y} V - \frac{\partial}{\partial y}\frac{\partial}{\partial x} V = 0, \end{aligned}$$

thus

$$\mathbf{b} = \operatorname{curl} \mathbf{a} = 0\,!$$

In other words, if the vector field $\mathbf{a}(x, y, z)$ is the gradient of the scalar field $V(x, y, z)$, *its curl will vanish*:

$$\operatorname{curl} \operatorname{grad} V = \langle \nabla \times \nabla V \rangle = \langle \nabla \times \nabla \rangle V = 0,$$

so the vectorial product of ∇ with itself vanishes just as if it were an ordinary vector!

Definition A vector field $\mathbf{a}$ of which the curl vanishes is called *conservative*, and the corresponding scalar field V, $\mathbf{a} = \operatorname{grad} V$, is called the *potential* of field $\mathbf{a}$.

We note immediately that, if

$$\mathbf{a}(x, y, z) = \operatorname{grad} V(x, y, z),$$

then also

$$\mathbf{a}(x, y, z) = \operatorname{grad} \left(V(x, y, z) + V_0 \right),$$

with V_0 an arbitrary constant, because

$$\operatorname{grad} V_0 = \mathbf{i} \frac{\partial V_0}{\partial x} + \mathbf{j} \frac{\partial V_0}{\partial y} + \mathbf{k} \frac{\partial V_0}{\partial z} = 0.$$

So the potential is *not uniquely defined*.

A.2.8 The Laplace operator

Assume a conservative field $\mathbf{a}$, so $\operatorname{curl} \mathbf{a} = 0$. Then we may write

$$\mathbf{a} = \operatorname{grad} V = \nabla V,$$

in which V is the potential.

Let us now express the *divergence* of field $\mathbf{a}$ into the potential:

$$\operatorname{div} \mathbf{a} = \langle \nabla \cdot \mathbf{a} \rangle = \langle \nabla \cdot \nabla V \rangle = \frac{\partial}{\partial x} \frac{\partial}{\partial x} V + \frac{\partial}{\partial y} \frac{\partial}{\partial y} V + \frac{\partial}{\partial z} \frac{\partial}{\partial z} V = \\ = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) V \stackrel{\text{def}}{=} \Delta V,$$

where we have introduced a new *differential operator*, the Delta operator invented by the French Pierre-Simon Laplace,

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} = \langle \nabla \cdot \nabla \rangle = \nabla^2.$$

When operating on the potential of a "source free" field — for example the gravitational potential in a vacuum or the electrostatic potential in an area of space free of electric charges — the result of this Delta, or Laplace, operator vanishes.

A.3 Integrals

A.3.1 The curve integral

We saw earlier that work ΔE can be written as the scalar product of force $\mathbf{F}$ and path $\Delta\mathbf{r}$:

$$\Delta E = \langle \mathbf{F} \cdot \Delta\mathbf{r} \rangle.$$

The differential form of this is

$$dE = \langle \mathbf{F} \cdot d\mathbf{r} \rangle,$$

from which one obtains the integral form, the *work integral*

$$\Delta E_{AB} = \int_A^B \langle \mathbf{F} \cdot d\mathbf{r} \rangle.$$

Here, the amount of work done by a body moving from point A to point B is computed by integrating $\langle \mathbf{F} \cdot d\mathbf{r} \rangle$ along the path AB.

If we parametrise the path according to arc length s, and the tangent vector to the path is called

$$\mathbf{t} \stackrel{\text{def}}{=} \frac{dx}{ds}\mathbf{i} + \frac{dy}{ds}\mathbf{j} + \frac{dz}{ds}\mathbf{k},$$

we may also write

$$\Delta E_{AB} = \int_A^B \langle \mathbf{F} \cdot \mathbf{t} \rangle \, ds,$$

the parametrised version of the integral.

A.3.2 The surface integral

Assume we are given again a vector field $\mathbf{a}$ and a surface in space S. Often, one seeks to integrate over surface S the *normal component* of a

vector field, the projection of $\mathbf{a}$ onto the normal vector of the surface, the vector perpendicular to the surface in the surface element dS.

Let the normal vector on the surface be $\mathbf{n}$. Then we must integrate

$$\iint_S \langle \mathbf{a} \cdot \mathbf{n} \rangle \, dS,$$

symbolically written

$$\iint_S \langle \mathbf{a} \cdot d\mathbf{S} \rangle,$$

in which the notation $d\mathbf{S}$ is called an *oriented surface element*. It is a vector pointing in the same direction as the normal vector $\mathbf{n}$.

Like a curve, a surface can also be *parametrised*. For example, the Earth's surface (assumed a sphere) can be parametrised by latitude ϕ and longitude λ: $\mathbf{r} = \mathbf{r}(\phi, \lambda)$. In this case we write as the surface element

$$dS = R^2 \cos\phi \, d\phi \, d\lambda,$$

in which $R^2 \cos\phi$ is *Jacobi's determinant* of the parameter pair (ϕ, λ). In this parametrisation, the integral is calculated as follows:

$$\iint_S \langle \mathbf{a} \cdot d\mathbf{S} \rangle = \iint_S \langle \mathbf{a} \cdot \mathbf{n} \rangle \, dS = \int_0^{2\pi} \int_{-\pi/2}^{+\pi/2} \langle \mathbf{a} \cdot \mathbf{n} \rangle \, R^2 \cos\phi \, d\phi \, d\lambda.$$

Other surfaces and parametrisations have other Jacobi's determinants. The determinant always represents the true area of a "parameter surface element" $d\phi \, d\lambda$ "in nature". For example, on the Earth's surface, a degree times degree patch is largest near the equator. In polar coordinates (ρ, θ) in the plane $(x = \rho\cos\theta, y = \rho\sin\theta)$, the determinant of Jacobi is ρ. In the ordinary (x, y) parametrisation in the plane, Jacobi's determinant is 1 and thus can be left out altogether.

A.3.3 The Stokes curl theorem

Let S be a surface in space (not necessarily flat) and ∂S its edge curve. Assume that the surface and its edge are well-behaved enough for all

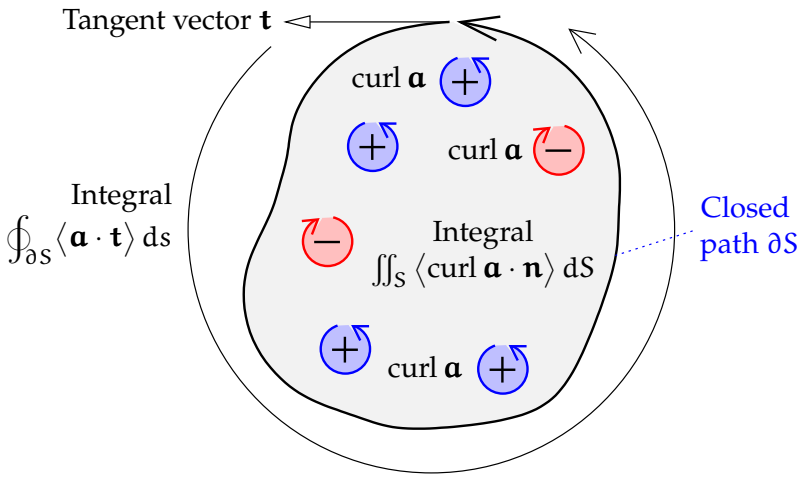


FIGURE A.6. The Stokes curl theorem.

necessary integrations and differentiations to be possible. Then (also called the Kelvin[1]–Stokes theorem):

$$\iint_S \langle \text{curl } \mathbf{a} \cdot d\mathbf{S} \rangle = \oint_{\partial S} \langle \mathbf{a} \cdot d\mathbf{r} \rangle,$$

with $\mathbf{r}$ the location vector of the edge curve. The parametrised form of the theorem is

$$\iint_S \langle \text{curl } \mathbf{a} \cdot \mathbf{n} \rangle \, dS = \oint_{\partial S} \langle \mathbf{a} \cdot \mathbf{t} \rangle \, ds,$$

with $\mathbf{n}$ is the normal to surface S and $\mathbf{t}$ the tangent vector of edge curve ∂S.

In words The surface integral of the curl of a vector field over a surface is the same as the closed path integral of the field around the edge of the surface.

Special case For a conservative vector field $\mathbf{a}$ it holds that $\text{curl } \mathbf{a} = 0$ everywhere. Then

$$\oint_{\partial S} \langle \mathbf{a} \cdot d\mathbf{r} \rangle = 0,$$

[1] William Thomson, Lord Kelvin PRS FRSE (1824–1907) was a British physicist, mathematician, engineer and inventor, and of course best known for the absolute temperature scale proposed by him in 1848.

so also

$$\int_{\substack{A\\ \text{path 1}}}^{B} \langle \mathbf{a} \cdot \mathrm{d}\mathbf{r} \rangle = \int_{\substack{A\\ \text{path 2}}}^{B} \langle \mathbf{a} \cdot \mathrm{d}\mathbf{r} \rangle .$$

Let $\mathbf{a}$ be the force vector of a field, like the acceleration, or force per unit mass, caused by the gravity field. Then this has the following interpretation:

The work integral from point A to point B does not depend on the path chosen. And the work done by a body transported around a closed path is zero.

This perhaps explains better the essence of a conservative force field. A conservative field can be represented as the *gradient of a potential*: $\mathbf{a} = \operatorname{grad} V$, in which V is the potential of the field. The Earth's gravity field $\mathbf{g}(x, y, z)$ is the gradient of the Earth's gravity potential or geopotential $W(x, y, z)$. At mean sea level — more precisely, at the geoid — the gravity *potential* is constant; the gravity *vector* $\mathbf{g}$ is everywhere perpendicular to the geoid.

A.3.4 The Gauss divergence theorem

Let $\mathcal{V}$ be a part of space, and $\partial\mathcal{V}$ its closed boundary, a union of surfaces. Assume again that both are mathematically well-behaved. Then the following theorem applies (Gauss):

$$\iiint_{\mathcal{V}} \operatorname{div} \mathbf{a} \, \mathrm{d}\mathcal{V} = \iint_{\partial\mathcal{V}} \langle \mathbf{a} \cdot \mathrm{d}\mathbf{S} \rangle = \iint_{\partial\mathcal{V}} \langle \mathbf{a} \cdot \mathbf{n} \rangle \, \mathrm{dS}.$$

In words What is created inside a body ("sources", divergence) must come out through its surfaces.

Usually, the orientation of surface $\partial\mathcal{V}$ is taken as positive on the outside: the normal vector $\mathbf{n}$ of the surface points outwards.

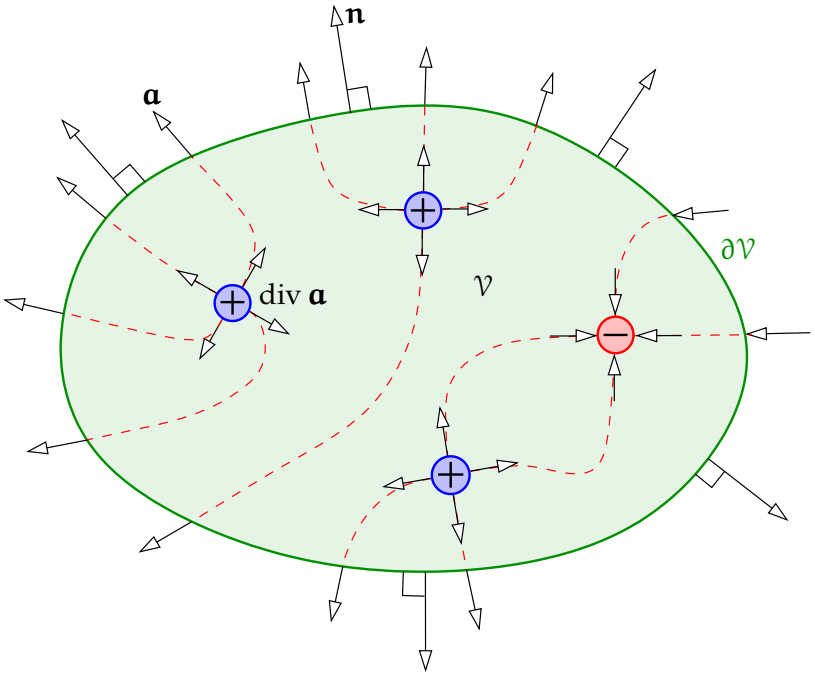


FIGURE A.7. The Gauss divergence theorem. $\mathbf{n}$ is the normal vector to the exterior surface. The Gauss divergence theorem can also be presented with the aid of (Michael Faraday's) *field lines*: a field line starts or terminates on an electric charge (a place where div $\mathbf{a} \neq 0$) or runs to infinity (through the surface ∂V).

A.4 The continuity of matter

An often-used equation in hydro- and aerodynamics is the *continuity equation*. This expresses that matter cannot just disappear or increase in amount. In the general case, the equation looks like this:

$$\operatorname{div}(\rho \boldsymbol{v}) + \frac{d}{dt}\rho = 0.$$

Here, the expression $\rho\boldsymbol{v}$ stands for *mass currents*, ρ is the matter density, $\boldsymbol{v}$ is the velocity of flow. The term $\operatorname{div}(\rho\boldsymbol{v})$ expresses how much more matter, in a unit of time, exits the volume element than enters it, per unit of volume. The second term again, the time derivative of the density ρ, stands for the change in the amount of matter inside the volume element

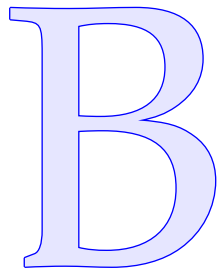
over time. The two terms must balance for the "matter accounting" to close.

If the moving fluid is incompressible, then ρ is constant:

$$\frac{d}{dt}\rho = 0 \quad \Longrightarrow \quad \operatorname{div}(\rho \boldsymbol{v}) = \rho \operatorname{div} \boldsymbol{v} = 0 \quad \Longrightarrow \quad \operatorname{div} \boldsymbol{v} = 0.$$

When talking about liquid or gas flow, it is good to be aware that the *vorticity* $\operatorname{curl} \boldsymbol{v}$ does *not* necessarily vanish — so, the flow is not necessarily eddy-free — so a potential V for which $\boldsymbol{v} = \operatorname{grad} V$ *does not necessarily exist*. In fact, turbulent flow is very common, and even for laminar flow, usually $\operatorname{curl} \boldsymbol{v} \neq 0$.

Function spaces



B.1 An abstract vector space

In an *abstract vector space* we may create a *basis*, with the help of which each vector in the space can be expressed as a linear combination of the basis vectors: for example, if the basis, in a concrete three-dimensional space, is $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$, we may write an arbitrary vector $\mathbf{r}$ in the form

kanta

$$\mathbf{r} = r_1\boldsymbol{e}_1 + r_2\boldsymbol{e}_2 + r_3\boldsymbol{e}_3 = \sum_{i=1}^{3} r_i\boldsymbol{e}_i.$$

Precisely because three basis vectors (not in the same plane) are always enough, we call the ordinary (Euclidean) space three-dimensional.

In a vector space one can define a *scalar product*, which is a linear mapping from two vectors to one number ("bilinear form"):

$$\langle \mathbf{r} \cdot \mathbf{s} \rangle.$$

Linearity means that

$$\langle (\alpha\mathbf{r}_1 + \beta\mathbf{r}_2) \cdot \mathbf{s} \rangle = \alpha \langle \mathbf{r}_1 \cdot \mathbf{s} \rangle + \beta \langle \mathbf{r}_2 \cdot \mathbf{s} \rangle \qquad \alpha, \beta \in \mathbb{R}$$

and commutativity that

vaihdan-
naisuus

$$\langle \mathbf{r} \cdot \mathbf{s} \rangle = \langle \mathbf{s} \cdot \mathbf{r} \rangle.$$

kantavektori If the basis vectors are *orthogonal* to each other, in other words, $\langle \boldsymbol{e}_i \cdot \boldsymbol{e}_j \rangle = 0$ if $i \neq j$, we may calculate the coefficients r_i in a simple way:

$$\mathbf{r} = \sum_{i=1}^{3} r_i \boldsymbol{e}_i, \qquad r_i = \frac{\langle \mathbf{r} \cdot \boldsymbol{e}_i \rangle}{\langle \boldsymbol{e}_i \cdot \boldsymbol{e}_i \rangle} = \frac{\langle \mathbf{r} \cdot \boldsymbol{e}_i \rangle}{\|\boldsymbol{e}_i\|^2}. \tag{B.1}$$

If, in addition,

$$\langle \boldsymbol{e}_i \cdot \boldsymbol{e}_i \rangle = \|\boldsymbol{e}_i\|^2 = 1, \qquad i \in \{1, 2, 3\},$$

in other words, the basis vectors are *orthonormal*, equation B.1 becomes simpler still:

$$\mathbf{r} = \sum_{i=1}^{3} r_i \boldsymbol{e}_i, \qquad r_i = \langle \mathbf{r} \cdot \boldsymbol{e}_i \rangle. \tag{B.2}$$

The quantity

$$\|\boldsymbol{e}_i\| \stackrel{\text{def}}{=} \sqrt{\langle \boldsymbol{e}_i \cdot \boldsymbol{e}_i \rangle}$$

is called the *norm* of the vector $\boldsymbol{e}_i$.

Unlike ordinary space, which is three-dimensional, a function space is an infinite-dimensional, *abstract* vector space, that nevertheless helps us to make certain abstract, but very useful fundamentals of function theory more concrete!

B.2 The Fourier function space

B.2.1 Description

Functions can also be considered elements in a vector space. If we define
[1] the scalar product of two functions f and g as the following integral[1]

$$\langle f \cdot g \rangle = \left\langle \overrightarrow{f} \cdot \overrightarrow{g} \right\rangle \stackrel{\text{def}}{=} \frac{1}{\pi} \int_0^{2\pi} f(x)\, g(x)\, dx, \tag{B.3}$$

it is easily verified that the above requirements for a scalar product are met.

kanta One basis in this vector space (a *function space*) is formed by the *Fourier*

[1] The arrows over the function designators f and g here are just psychology: the functions are really "vectors". These arrows are not normally used.

basis functions,

$$\begin{aligned} e_0 &= \frac{1}{2}\sqrt{2}, \\ e_k &= \cos kx, \qquad k = 1, 2, 3, \ldots, \\ e_{-k} &= \sin kx, \qquad k = 1, 2, 3, \ldots. \end{aligned} \tag{B.4}$$

This basis is *orthonormal* (proof: exercise). It is also a *complete* basis, which we shall not prove. As the number of basis vectors is countably infinite, we say that this function space is infinitely dimensional.

Now every function $f(x)$ meeting certain conditions can be expanded in the way of equation B.2, as follows:

$$f(x) = \frac{1}{2}a_0\sqrt{2} + \sum_{k=1}^{\infty} \left(a_k \cos kx + b_k \sin kx\right),$$

— the familiar Fourier-series expansion — in which the coefficients are

$$\begin{aligned} a_0 = \langle f \cdot e_0 \rangle &= \frac{1}{2\pi}\sqrt{2}\int_0^{2\pi} f(x)\,dx = \sqrt{2}\cdot\overline{f(x)}, \\ a_k = \langle f \cdot e_k \rangle &= \frac{1}{\pi}\int_0^{2\pi} f(x)\cos kx\,dx, \qquad k = 1, 2, 3, \ldots, \\ b_k = \langle f \cdot e_{-k} \rangle &= \frac{1}{\pi}\int_0^{2\pi} f(x)\sin kx\,dx, \qquad k = 1, 2, 3, \ldots. \end{aligned}$$

This is the familiar way in which the coefficients of a Fourier series are calculated.

B.2.2 Example

As an example of Fourier analysis, we may take a step function on the interval $[0, 2\pi)$:

$$f(x) = \begin{cases} 0 & x \in [0, \pi), \\ 1 & x \in [\pi, 2\pi). \end{cases}$$

We can calculate the Fourier coefficients of this function as follows:

$$
\begin{aligned}
a_0 &= \frac{1}{2\pi}\sqrt{2}\cdot\int_0^{2\pi} f(x)\,dx = \frac{1}{2\pi}\sqrt{2}\cdot\pi = \frac{1}{2}\sqrt{2},\\
a_k &= \frac{1}{\pi}\int_0^{2\pi} f(x)\cos kx\,dx = \frac{1}{\pi}\int_\pi^{2\pi}\cos kx\,dx =\\
&= \frac{1}{\pi}\left[\frac{1}{k}\sin kx\right]_\pi^{2\pi} = \frac{1}{k\pi}\left(\sin 2k\pi - \sin k\pi\right) = 0,\\
b_k &= \frac{1}{\pi}\int_0^{2\pi} f(x)\sin kx\,dx = \frac{1}{\pi}\int_\pi^{2\pi}\sin kx\,dx =\\
&= \frac{1}{\pi}\left[-\frac{1}{k}\cos kx\right]_\pi^{2\pi} = \frac{1}{k\pi}\left(\cos k\pi - \cos 2k\pi\right) =\\
&= \frac{1}{k\pi}\left((-1)^k - 1\right) = \begin{cases} 0 & \text{if } k \text{ even},\\ -\dfrac{2}{k\pi} & \text{if } k \text{ odd}.\end{cases}
\end{aligned}
$$

In numbers: $a_0 = \frac{1}{2}\sqrt{2} = 0.707\,10\ldots$, $b_1 = -2/\pi = -0.636\,62\ldots$, $b_3 = -2/3\pi = -0.212\,20\ldots$, $b_5 = -0.127\,32\ldots$, and so forth. The expansion now becomes

$$f(x) = \tfrac{1}{2}\sqrt{2}\,a_0 + \sum_{k=1}^{\infty} b_k \sin kx = \tfrac{1}{2} - \frac{2}{\pi}\sum_{\substack{k=1\\ \text{odd}}}^{\infty}\frac{1}{k}\sin kx.$$

We see that it only contains sines, no cosines. This is a consequence of the function's symmetry properties.

In figure B.1 we show truncated expansions for this function:

$$f^{(K)}(x) \stackrel{\text{def}}{=} \tfrac{1}{2}a_0\sqrt{2} + \sum_{k=1}^{K} b_k \sin kx = \tfrac{1}{2} - \frac{2}{\pi}\sum_{\substack{k=1\\ \text{odd}}}^{K}\frac{1}{k}\sin kx, \tag{B.5}$$

with K the truncation parameter.

B.2.3 Convergence

suppeneminen The Fourier expansion converges in the square integral sense: if we define the truncated expansion

$$f^{(K)}(x) \stackrel{\text{def}}{=} \tfrac{1}{2}a_0\sqrt{2} + \sum_{k=1}^{K}\left(a_k\cos kx + b_k\sin kx\right),$$

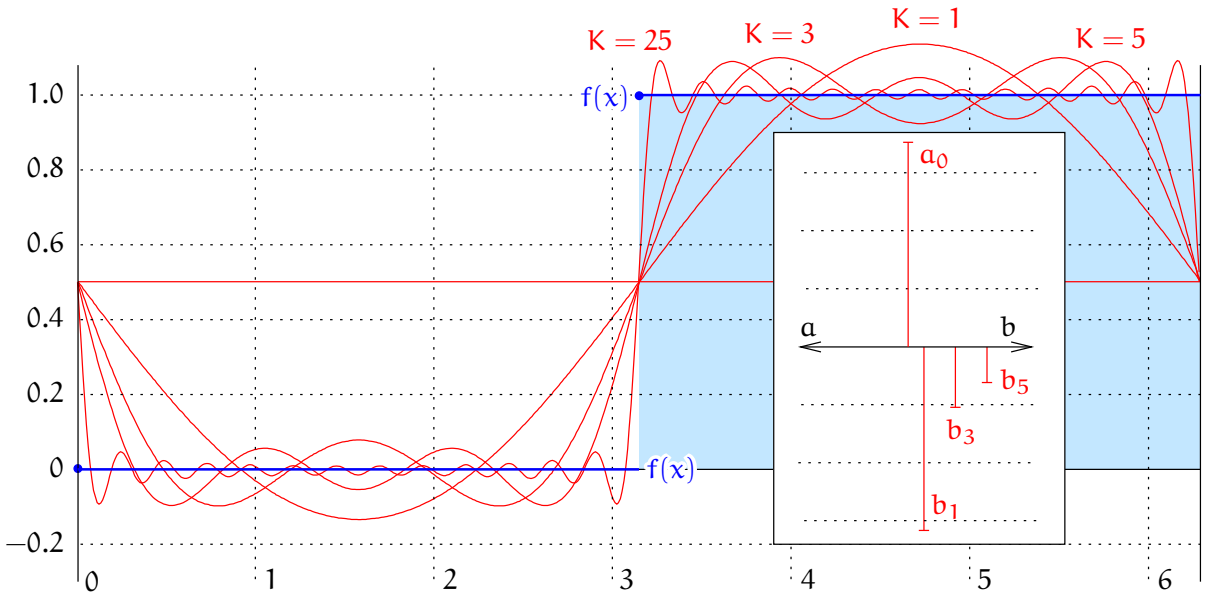


FIGURE B.1. Fourier analysis on a step function. Plotted are the truncated Fourier expansions $f^{(K)}(x)$, equation B.5, for values of K of 1, 3, 5, and 25. The inset gives the spectrum of the function.

then

$$\lim_{K\to\infty} \frac{1}{\pi} \int_0^{2\pi} \left(f^{(K)}(x) - f(x)\right)^2 \, dx = 0.$$

This does *not* mean that for an arbitrarily small value ε it holds that $\left|f^{(K)}(x) - f(x)\right| < \varepsilon$ for every $x \in [0, 2\pi)$, when $K \to \infty$. Looking at figure B.1, there will always remain a small neighbourhood of $x = \pi$ within which there will be points $x' \neq \pi$ where the absolute difference $\left|f^{(K)}(x') - f(x')\right| > 0.1$ (or any other positive bound < 0.5), even for arbitrarily large values of K. We say that the Fourier expansion is *convergent*, but not *uniformly* convergent.

tasainen suppeneminen

The Fourier expansion converges pointwise "almost everywhere" in $x \in [0, 2\pi)$: at all points except for the two special points $x = 0$ and $x = \pi$. By defining $f(0) = f(\pi) \stackrel{\text{def}}{=} 0.5$, the expansion is made pointwise convergent everywhere.

pisteittäinen suppeneminen

Also, note the "shoulder" of the expansion, even for $K = 25$. This

shoulder will get narrower for higher K, but not any lower, remaining at approximately 0.09. This is known as the Gibbs phenomenon.

Gibbsin ilmiö

B.3 Sturm–Liouville differential equations

B.3.1 The eigenvalue problem

In an abstract vector space we may formulate an eigenvalue problem: given a linear operator (mapping) L, we may write

$$L\mathbf{x} - \lambda\mathbf{x} = 0,$$

where the problem consists of determining the *eigenvalues* λ for which one or more solutions or *eigenvectors* $\mathbf{x}$ exist.

ortonormaali kanta

In a concrete n-dimensional vector space in which there is an orthonormal basis $\{\boldsymbol{e}_i, i = 1, \ldots, n\}$ we may write the vector

$$\mathbf{x} = \sum_{i=1}^{n} x_i \boldsymbol{e}_i,$$

and, thanks to linearity,

$$L\mathbf{x} = L\left(\sum_{i=1}^{n} x_i \boldsymbol{e}_i\right) = \sum_{i=1}^{n} x_i \cdot L\boldsymbol{e}_i.$$

On the other hand, we may write n different vectors $L\boldsymbol{e}_i$ on the basis $\{\boldsymbol{e}_j\}$ in the following way:

$$L\boldsymbol{e}_i = \sum_{j=1}^{n} a_{ij} \boldsymbol{e}_j, \qquad i = 1, \ldots, n.$$

This defines the coefficients a_{ij}, which may be collected into a size $n \times n$ matrix A.

Now substitution yields

$$L\mathbf{x} = \sum_{i=1}^{n} x_i \cdot \sum_{j=1}^{n} a_{ij} \boldsymbol{e}_j = \sum_{j=1}^{n} \left(\sum_{i=1}^{n} a_{ij} x_i\right) \boldsymbol{e}_j. \tag{B.6}$$

Also

$$\lambda \mathbf{x} = \lambda \sum_{i=1}^{n} x_i \mathbf{e}_i = \sum_{j=1}^{n} (\lambda x_j)\, \mathbf{e}_j. \tag{B.7}$$

By combining equations B.6 and B.7, of which all coefficients must be identical, we obtain

$$\sum_{i=1}^{n} a_{ij} x_i - \lambda x_j = 0, \qquad j = 1, \ldots, n,$$

or, as a matric equation,

$$A\overline{\mathbf{x}} - \lambda \overline{\mathbf{x}} = 0, \tag{B.8}$$

in which A is a matrix consisting of the coefficients a_{ij}, and $\overline{\mathbf{x}}$ a column vector consisting of the coefficients x_i: $\overline{\mathbf{x}} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^T$.

Of course equation B.8 also represents an eigenvalue problem, but now in the linear vector space $\mathbb{R}^n$ consisting of all *coefficient vectors* $\overline{\mathbf{x}}$. Every $\overline{\mathbf{x}}$ is the numerical representation of a vector $\mathbf{x}$ on the chosen basis $\{\mathbf{e}_i\}$. Matrix A is again the numerical representation of operator L on the same basis.[2]

B.3.2 A self-adjoint operator

Let L be a linear operator in a vector space where there exists a *scalar product*: a bilinear form $\langle \mathbf{x} \cdot \mathbf{y} \rangle$ which is symmetric or commutative.

Then L is *self-adjoint*, if for each pair of vectors $\mathbf{x}, \mathbf{y}$ it holds that

$$\langle \mathbf{x} \cdot L\mathbf{y} \rangle = \langle L\mathbf{x} \cdot \mathbf{y} \rangle.$$

If the corresponding matrix A is self-adjoint, that means that

$$\langle \overline{\mathbf{x}} \cdot A\overline{\mathbf{y}} \rangle = \langle A\overline{\mathbf{x}} \cdot \overline{\mathbf{y}} \rangle,$$

[2] An advantage of the numerical representations is of course that one can actually calculate with them.

or

$$\sum_{i=1}^{n} x_i \left(\sum_{j=1}^{n} a_{ij} y_j \right) = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} a_{ij} x_j \right) y_i,$$

which is trivially true if

$$a_{ij} = a_{ji}, \qquad i, j \in 1, \ldots, n, \qquad \text{or } A = A^{\mathsf{T}}.$$

In other words,

A symmetric matrix is a self-adjoint operator.

From linear algebra it is undoubtedly familiar that the eigenvectors $\mathbf{x}_p, \mathbf{x}_q$ belonging to different eigenvalues $\lambda_p \neq \lambda_q$ of a symmetric, size $n \times n$ matrix *are mutually orthogonal*: $\mathbf{x}_p \perp \mathbf{x}_q$. If all eigenvalues λ_p, $p = 1, \ldots, n$ are different, then the eigenvectors $\mathbf{x}_p$, $p = 1, \ldots, n$ will constitute a *complete orthogonal basis*[3] in the vector space $\mathbb{R}^n$.

täydellinen ortogonaali kanta

3

The proof is not hard. We start from the equation for the eigenvalue problem for eigenvectors and -values $\mathbf{x}_p, \lambda_p$:

$$L\mathbf{x}_p = \lambda_p \mathbf{x}_p,$$

and multiply from the left by vector $\mathbf{x}_q$:

$$\langle \mathbf{x}_q \cdot L\mathbf{x}_p \rangle = \lambda_p \langle \mathbf{x}_q \cdot \mathbf{x}_p \rangle.$$

Similarly for eigenvectors and -values $\mathbf{x}_q, \lambda_q$ multiplied from the left by vector $\mathbf{x}_p$:

$$\langle \mathbf{x}_p \cdot L\mathbf{x}_q \rangle = \lambda_q \langle \mathbf{x}_p \cdot \mathbf{x}_q \rangle.$$

If L is self-adjoint, then

$$\langle \mathbf{x}_q \cdot L\mathbf{x}_p \rangle = \langle L\mathbf{x}_q \cdot \mathbf{x}_p \rangle = \langle \mathbf{x}_p \cdot L\mathbf{x}_q \rangle \implies \lambda_p \langle \mathbf{x}_q \cdot \mathbf{x}_p \rangle = \lambda_q \langle \mathbf{x}_p \cdot \mathbf{x}_q \rangle.$$

---

[3] Actually the eigenvectors may be arbitrarily re-scaled: if $\mathbf{x}$ is an eigenvector, then also $\mathbf{e} \stackrel{\text{def}}{=} \mathbf{x}/\|\mathbf{x}\|$ is. Thus we obtain an *orthonormal basis*.

It follows that

$$(\lambda_p - \lambda_q)\left\langle \mathbf{x}_p \cdot \mathbf{x}_q \right\rangle = 0.$$

Remember that the scalar product is symmetric. If $\lambda_p \neq \lambda_q$, we thus must have $\left\langle \mathbf{x}_p \cdot \mathbf{x}_q \right\rangle = 0$, or $\mathbf{x}_p \perp \mathbf{x}_q$, what was to be proven.

Example *The variance matrix of location in the plane.* The variance matrix of the co-ordinates of point P in the plane is

$$\mathrm{Var}\{\overline{\mathbf{x}}_P\} = \mathrm{Var}\left\{\begin{bmatrix} \underline{x}_P \\ \underline{y}_P \end{bmatrix}\right\} = \Sigma_{PP} = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix},$$

a *symmetric* matrix. Here, σ_x^2 and σ_y^2 are the variances, or squares of the mean errors, of the x and y co-ordinates, and σ_{xy} is the covariance between the co-ordinates.

The eigenvalues of this matrix Σ_{PP} are the solutions of the *characteristic equation*

$$\det(\Sigma_{PP} - \lambda I) = \det\begin{bmatrix} \sigma_x^2 - \lambda & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 - \lambda \end{bmatrix} = 0,$$

or

$$(\sigma_x^2 - \lambda)(\sigma_y^2 - \lambda) - \sigma_{xy}^2 = 0.$$

This yields

$$\lambda_{1,2} = \tfrac{1}{2}(\sigma_x^2 + \sigma_y^2) \pm \tfrac{1}{2}\sqrt{(\sigma_x^2 + \sigma_y^2)^2 - 4(\sigma_x^2\sigma_y^2 - \sigma_{xy}^2)} = \\ = \tfrac{1}{2}(\sigma_x^2 + \sigma_y^2) \pm \tfrac{1}{2}\sqrt{(\sigma_x^2 - \sigma_y^2)^2 + 4\sigma_{xy}^2}.$$

The visual presentation of the variance matrix is the *variance* or *error ellipse*. The semi-lengths of its principal axes are $\sqrt{\lambda_1}$ and $\sqrt{\lambda_2}$, and the directions of the principal axes are the eigenvectors of Σ_{PP}: $\overline{\mathbf{x}}_1$ and $\overline{\mathbf{x}}_2$, mutually orthogonal. If the co-ordinate axes are turned into the directions of $\overline{\mathbf{x}}_{1,2}$, the matrix Σ_{PP} will assume the form

$$\Sigma'_{PP} = \begin{bmatrix} \sigma_{x'}^2 & 0 \\ 0 & \sigma_{y'}^2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}.$$

The sum of the eigenvalues (and the *trace* of the matrix), $\lambda_1 + \lambda_2 = \sigma_x^2 + \sigma_y^2$, is an *invariant* called the *point variance*.

B.3.3 Self-adjoint differential equations

A function space also features self-adjoint or "symmetric" differential equations. In fact, the most famous equations of physics are of this type.

Take a good look at, for example, the oscillation equation, in which $x(t)$ is the position as a function of time:

$$\frac{d^2}{dt^2}x(t) + \omega^2 x(t) = 0. \tag{B.9}$$

The solution has the general form (α amplitude, ϕ phase constant)

$$x(t) = \alpha \sin(\omega t - \phi).$$

On the interval $t \in [0, T]$ we require *periodicity*:

$$x(0) = x(T), \qquad \frac{d}{dt}x\Big|_{x=0} = \frac{d}{dt}x\Big|_{x=T}.$$

These *boundary conditions* are an essential part of being self-adjoint. Then, a solution is found only for certain values of ω — *quantisation*.

Equation B.9 is an *eigenvalue problem*, form-wise:

$$Lx + \omega^2 x = 0,$$

in which the operator is

$$L = \frac{d^2}{dt^2}.$$

We first show that this operator is on the interval $[0, T]$ self-adjoint. If the scalar product is defined as follows:

$$\langle \vec{x} \cdot \vec{y} \rangle \stackrel{\text{def}}{=} \int_0^T x(t)\, y(t)\, dt,$$

it holds that (integration by parts):

$$\left\langle \vec{x} \cdot L\vec{y} \right\rangle = \int_0^T x(t) \frac{d^2y(t)}{dt^2} dt = \left[x(t) \frac{dy(t)}{dt} \right]_0^T - \int_0^T \frac{dx(t)}{dt} \frac{dy(t)}{dt} dt,$$
$$\left\langle L\vec{x} \cdot \vec{y} \right\rangle = \int_0^T \frac{d^2x(t)}{dt^2} y(t) dt = \left[\frac{dx(t)}{dt} y(t) \right]_0^T - \int_0^T \frac{dx(t)}{dt} \frac{dy(t)}{dt} dt.$$

As, on the right-hand side, the first terms vanish and the second terms are identical, it follows that

$$\left\langle \vec{x} \cdot L\vec{y} \right\rangle = \left\langle L\vec{x} \cdot \vec{y} \right\rangle,$$

which was to be proven.

Self-adjoint operators have eigenvalues and eigenvectors, in this case *functions*, that are mutually orthogonal for different values of ω.[4] For the oscillation equation with the above periodicity conditions they are just the solution functions

$$\sin(\omega_k t - \phi) = \sin\left(\frac{2\pi k}{T} t - \phi \right), \tag{B.10}$$

in which the frequency

$$\omega_k = \frac{2\pi k}{T}$$

is quantised by a "quantum number" $k \in \mathbb{N}$.

If we let $T \to \infty$, the frequencies ω_k get closer and closer to each other, and in the end morph into a continuum.

In physics there is a broad class of differential equations that are self-adjoint in some function space. The class is known as "Sturm[5]–Liouville[6]

[4] In fact, for the same value ω_k there exist two mutually orthogonal periodic solutions,

$$\sin \omega_k t = \sin \frac{2\pi k t}{T}, \qquad \cos \omega_k t = \cos \frac{2\pi k t}{T}.$$

Any linear combination of these is a valid solution as well, and is of the general form B.10.

[5] Jacques Charles François Sturm FRS FAS (1803–1855) was a French mathematician, one of the 72 names engraved on the Eiffel Tower. Eiffel Tower, 72 names.

type problems". It includes the oscillation equation, Legendre's equation, Bessel's equation, and many more. Every one of them generates, in a natural way, its own set of mutually orthogonal functions that serve as
kantafunktio the basis functions for the general solution of many partial differential equations.

B.4 Legendre polynomials

The ordinary Legendre polynomials $P_n(t)$ also constitute a basis in a function space, with the scalar product definition

$$\left\langle \vec{f} \cdot \vec{g} \right\rangle \stackrel{\text{def}}{=} \int_{-1}^{+1} f(t)\, g(t)\, dt.$$

They do not however constitute an ortho*normal* basis, but only an
ortogonaali kanta ortho*gonal* one:

$$\|P_n\|^2 = \langle P_n \cdot P_n \rangle = \int_{-1}^{+1} P_n^2(t)\, dt = \frac{2}{2n+1}.$$

B.5 Spherical harmonics

On the surface of a sphere, all functions can also be considered elements of a function space. Every function meeting certain well-behavedness requirements — like integrability — is an element. The functions

$$\begin{aligned} R_{nm}(\phi,\lambda) &= P_{nm}(\sin\phi)\cos m\lambda, \qquad n = 0,1,2,\ldots, \quad m = 0,\ldots,n, \\ S_{nm}(\phi,\lambda) &= P_{nm}(\sin\phi)\sin m\lambda, \qquad n = 0,1,2,\ldots, \quad m = 1,\ldots,n, \end{aligned}$$

together form a *complete basis* for this vector space in such a way that every function can be written as an — if necessary infinite — linear combination of these basis functions. The situation is analogous to three-dimensional space, where a complete basis consists of three vectors not in the same plane.

---

[6] Joseph Liouville FRS FRSE FAS (1809–1882) was a French mathematician.

An alternative, more compact way of writing this is

$$Y_{nm}(\phi,\lambda) = \begin{cases} P_{nm}(\sin\phi)\cos m\lambda & \text{if } m \geqslant 0, \\ P_{n|m|}(\sin\phi)\sin|m|\lambda & \text{if } m < 0, \end{cases}$$

for values $n = 0, 1, 2, \ldots$, $m = -n, \ldots, n$.

In this function space, a *scalar product* is defined:

$$\left\langle \vec{f} \cdot \vec{g} \right\rangle = \frac{1}{4\pi} \iint_\sigma f(\phi,\lambda)\, g(\phi,\lambda)\, d\sigma,$$

in which σ is the surface of the unit sphere ("directional sphere", or even "celestial sphere"), $d\sigma = \cos\phi\, d\phi\, d\lambda$ is a surface element of this sphere, and $\cos\phi$ is the determinant of Jacobi of the co-ordinates (ϕ,λ).

According to this definition, we can show that two different functions, Y_{nm} and $Y_{n'm'}$, are orthogonal with respect to each other:

$$\langle Y_{nm} \cdot Y_{n'm'} \rangle = \frac{1}{4\pi} \iint_\sigma Y_{nm}(\phi,\lambda)\, Y_{n'm'}(\phi,\lambda)\, d\sigma = 0$$

if $n \neq n'$ or $m \neq m'$.

The basis $\{Y_{nm}, n = 0, 1, 2, \ldots, m = -n, \ldots, n\}$ is ortho*gonal* but not ortho*normal*: the lengths of the vectors differ from unity.

$$\|Y_{nm}\|^2 = \langle Y_{nm} \cdot Y_{nm} \rangle =$$
$$= \frac{1}{4\pi} \iint_\sigma Y_{nm}^2(\phi,\lambda)\, d\sigma = \begin{cases} \dfrac{1}{2n+1} & \text{if } m = 0, \\ \dfrac{1}{2(2n+1)} \dfrac{(n+|m|)!}{(n-|m|)!} & \text{if } m \neq 0, \end{cases}$$

see Heiskanen and Moritz (1967, equation 1-69). Proving this orthogonality is not straightforward.

If we now divide the functions Y_{nm} (or, equivalently, R_{nm}, S_{nm}) by the square roots of the above factors, we obtain the *fully normalised* surface spherical harmonics $\overline{Y}_{nm}$, for which it holds that

pinta-
pallofunktio

$$\left\|\overline{Y}_{nm}\right\|^2 = \frac{1}{4\pi} \iint_\sigma \overline{Y}_{nm}^2(\phi,\lambda)\, d\sigma = 1.$$

With those it is again easy to calculate the coefficients $\overline{f}_{nm}$ of a given general function on the sphere $f(\phi,\lambda)$ (the overline means that these are fully normalised coefficients):

$$\overline{f}_{nm} = \langle f \cdot \overline{Y}_{nm} \rangle = \frac{1}{4\pi} \iint_\sigma f(\phi,\lambda)\, \overline{Y}_{nm}(\phi,\lambda)\, d\sigma. \tag{B.11}$$

This is, in the geometric analogy, a straightforward *projection* onto the unit vectors of the basis.

In the above integral, $f(\phi,\lambda)$ is the function f *on the Earth's surface*: if the radius of the spherical Earth is R, then $f(\phi,\lambda) = f(\phi,\lambda,R)$.

The fully normalised equation corresponding to expansion 2.12 is

$$V(\phi,\lambda,r) = \sum_{n=0}^{\infty} \frac{1}{r^{n+1}} \sum_{m=0}^{n} \overline{P}_{nm}(\sin\phi) \left(\overline{a}_{nm} \cos m\lambda + \overline{b}_{nm} \sin m\lambda\right).$$

We may also write

$$\overline{Y}_{nm}(\phi,\lambda) = \begin{cases} \overline{P}_{nm}(\sin\phi)\cos m\lambda & \text{if } m \geqslant 0, \\ \overline{P}_{n|m|}(\sin\phi)\sin|m|\lambda & \text{if } m < 0, \end{cases}$$

which corresponds to the definition of the fully normalised Legendre functions:

$$\begin{aligned} \overline{P}_{n0}(\sin\phi) &= \sqrt{2n+1}\, P_{n0}(\sin\phi), \\ \overline{P}_{nm}(\sin\phi) &= \sqrt{2(2n+1)\frac{(n-m)!}{(n+m)!}}\, P_{nm}(\sin\phi), \qquad m > 0. \end{aligned}$$

Now, the above equation for the potential becomes

$$V(\phi,\lambda,r) = \sum_{n=0}^{\infty} \frac{1}{r^{n+1}} \sum_{m=-n}^{n} \overline{v}_{nm} \overline{Y}_{nm}(\phi,\lambda),$$

in which

$$\overline{v}_{nm} = \begin{cases} \overline{a}_{nm} & \text{if } m \geqslant 0, \\ \overline{b}_{n|m|} & \text{if } m < 0. \end{cases}$$

On the sphere $r = R$ this becomes

$$V(\phi, \lambda, R) = \sum_{n=0}^{\infty} \frac{1}{R^{n+1}} \sum_{m=-n}^{n} \bar{v}_{nm} \overline{Y}_{nm}(\phi, \lambda),$$

from which by orthogonal projection (equation B.11) follows

$$\bar{v}_{nm} = R^{n+1} \left\langle V \cdot \overline{Y}_{nm} \right\rangle = \frac{R^{n+1}}{4\pi} \iint_{\sigma} V(\phi, \lambda, R)\, \overline{Y}_{nm}(\phi, \lambda)\, d\sigma$$

or

$$\bar{a}_{nm} = \frac{R^{n+1}}{4\pi} \iint_{\sigma} V(\phi, \lambda, R)\, \overline{P}_{nm}(\phi, \lambda) \cos m\lambda\, d\sigma,$$
$$\bar{b}_{nm} = \frac{R^{n+1}}{4\pi} \iint_{\sigma} V(\phi, \lambda, R)\, \overline{P}_{nm}(\phi, \lambda) \sin m\lambda\, d\sigma.$$

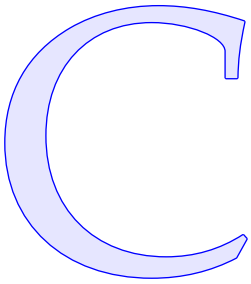
Self-test questions

1. The identity $\langle \mathbf{r} \cdot \mathbf{s} \rangle = \langle \mathbf{s} \cdot \mathbf{r} \rangle$, for two elements $\mathbf{r}$ and $\mathbf{s}$ of a vector space, expresses the property of linearity | commutativity | associativity.

Exercise B−1: Orthonormality of the Fourier basis functions

Show the orthonormality of the Fourier basis functions, equation B.4 by deriving their scalar products by equation B.3.

Why does FFT work?



FFT is a factorisation method for computing the discrete Fourier transform that spectacularly reduces the number of calculations needed and speeds up the calculation. It requires the number of data grid points to be a factorisable number.

There are alternatives in choosing precisely which FFT method to use. The fastest FFT requires a grid the number of points of which is a power of 2. The size of the grid is then $2^n \times 2^m$. Alternative, "mixed-radix" methods may also be considered and perform well if the grid size is something like 360×480, for example $N = 360 = 2 \times 2 \times 2 \times 3 \times 3 \times 5$. If the grid size is a prime number, FFT is no better than the ordinary discrete Fourier transform.

If the function $f(x)$ is given on the interval $x \in [0, L)$, on an equispaced grid, $x_k = {}^{kL}/_{N}$, as values $f_k = f(x_k)$, $k = 0, \ldots, N-1$, the discrete Fourier transform in one dimension is

$$\mathcal{F}\{f(x)\} = F(\tilde{\nu}),$$

in which

$$F(\tilde{\nu}_j) = \frac{1}{N} \sum_{k=0}^{N-1} f(x_k) \exp\left(-2\pi i \frac{jk}{N}\right), \qquad j = 0, \ldots, N-1. \tag{C.1}$$

The frequency argument, spatial frequency or wave number, $\tilde{\nu}_j = {}^{j}/_{L}$, $j = 0, \ldots, N-1$ is defined on the interval[1] $\left[0, {}^{(N-1)}/_{L}\right]$. i is the [1]

imaginary unit: $i^2 = -1$. We use $\exp(x)$ to denote e^x.

Correspondingly, the inverse discrete Fourier transform,

$$\mathcal{F}^{-1}\{F(\tilde{\nu})\} = f(x),$$

is

$$f(x_k) = \sum_{j=0}^{N-1} F(\tilde{\nu}_j) \exp\left(2\pi i \frac{jk}{N}\right), \qquad k = 0, \ldots, N-1. \tag{C.2}$$

FFT is just a brutally efficient way of computing both these equations C.1 and C.2. A brute-force calculation of these formulas requires an order of N^2 "standard operations", each of them a single multiplication plus a single addition or subtraction. If N is even, we may write

$$\begin{aligned}
F(\tilde{\nu}_j) &= \frac{1}{N}\left(\sum_{k=0}^{\frac{1}{2}N-1} f_k \exp\left(-2\pi i \frac{jk}{N}\right) + \sum_{k=\frac{1}{2}N}^{N-1} f_k \exp\left(-2\pi i \frac{jk}{N}\right)\right) = \\
&= \frac{1}{N}\left(\sum_{k=0}^{\frac{1}{2}N-1} f_k \exp\left(-2\pi i \frac{jk}{N}\right) + \exp\left(-2\pi i j \frac{\frac{1}{2}N}{N}\right) \sum_{k'=0}^{\frac{1}{2}N-1} f_{k'+\frac{1}{2}N} \exp\left(-2\pi i \frac{jk'}{N}\right)\right) = \\
&= \frac{1}{N}\left(\sum_{k=0}^{\frac{1}{2}N-1} f_k \exp\left(-2\pi i \frac{jk}{N}\right) + \exp(-\pi i j) \sum_{k=0}^{\frac{1}{2}N-1} f_{k+\frac{1}{2}N} \exp\left(-2\pi i \frac{jk}{N}\right)\right) = \\
&= \frac{1}{N} \sum_{k=0}^{\frac{1}{2}N-1} \left[f_k \pm f_{k+\frac{1}{2}N}\right] \exp\left(-2\pi i \frac{jk}{N}\right), \qquad \left\{\begin{matrix} + \text{ if } j \text{ even} \\ - \text{ if } j \text{ odd}\end{matrix}\right\}
\end{aligned} \tag{C.3}$$

the computation of which sum requires only $N \cdot \frac{1}{2}N$ multiplications and additions/subtractions, not counting pre-calculations.

Here we used Euler's identity $\exp(-\pi i) = -1$, so $e^{-\pi i j} = \left(e^{-\pi i}\right)^j = (-1)^j$, either $+1$ or -1.[2] The expression in square brackets, for each k

[2]

[1] Alternatively, the interval of definition can be chosen as $\left[-\frac{1}{2}\,{}^{N}\!/_{L}\,,\ (\frac{1}{2}N-1)\,/_{L}\right]$. This is done by mapping $\tilde{\nu}_j \to \tilde{\nu}_j - {}^{N}\!/_{L}$, or $j \to j-N$, for $j > \frac{1}{2}N-1$. This has the merit of placing the frequency zero in the middle. It does not materially change anything, as it simply multiplies $F(\tilde{\nu}_j)$ with unity: $\exp(-2\pi i\, {}^{Nk}\!/_{N}) = \exp(-2\pi i k) = 1$, the periodicity property of the discrete Fourier transform.

value $k = 0, 1, \ldots, \frac{1}{2}N - 1$, is either a summation, for even values of j, or a subtraction, for odd values of j. In total, $\frac{1}{2}N$ sums and $\frac{1}{2}N$ differences are pre-calculated. Also the exp expressions are pre-calculated into a lookup table.

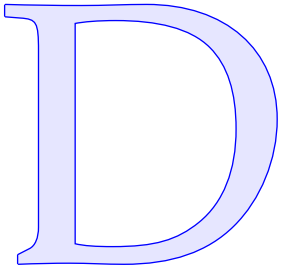
Altogether some $\frac{1}{2}N^2$ standard operations are needed, half the original number.

Equation C.3 is itself recognised as a Fourier series, but the number of support points is only $\frac{1}{2}N$ instead of N. If $\frac{1}{2}N$ is also even, we may repeat the above trick, resulting in an expression requiring only an order of $\frac{1}{4}N^2$ operations. Lather, rinse, repeat, and the number of operations becomes $\frac{1}{8}N^2$, $\frac{1}{16}N^2$, $\frac{1}{32}N^2$, etc.... A more precise analysis shows that if N is a power of 2, the whole discrete Fourier transform may be computed in order $N \cdot {}^2\log N$ operations!

In the literature, smart algorithms are found implementing the method described, for example fftw ("Fastest Fourier Transform in the West", FFTW Home Page; Frigo and Johnson, 2005).

[2] These values are called the "twiddle factors".

Helmert condensation



In order to derive the equation for Helmert condensation, we first derive the equation for the potential of the topography:

$$V_{\text{top}}(\phi,\lambda,r) = G\iiint_{\text{top}} \frac{\rho(\phi',\lambda',r')}{\ell(\psi,r,r')}\,d\mathcal{V}' \approx G\rho\iiint_{\text{top}} \frac{1}{\ell(\psi,r,r')}\,d\mathcal{V}',$$

in which ψ is the geocentric angular distance between the evaluation point (ϕ,λ,r) and the data point (ϕ',λ',r'). We assume a standard density ρ.

We similarly derive the equation for the potential of the condensation layer:

$$V_{\text{cond}}(\phi,\lambda,r) = G\rho\iiint_{\text{cond}} \frac{1}{\ell(\psi,r,R)}\,d\mathcal{V}'.$$

We integrate in spherical co-ordinates:

$$\begin{aligned}
\iiint_{\text{top}} \frac{1}{\ell(\psi,r,r')}\,d\mathcal{V}' &= \int_\sigma \int_R^{R+H(\phi',\lambda')} \frac{1}{\ell(\psi,r,r')}\,(r')^2\,dr'\,d\sigma',\\
\iiint_{\text{cond}} \frac{1}{\ell(\psi,r,R)}\,d\mathcal{V}' &= \int_\sigma \frac{1}{\ell(\psi,r,R)} \int_R^{R+H(\phi',\lambda')} (r')^2\,dr'\,d\sigma' =\\
&= R^2\int_\sigma \frac{H(\phi',\lambda')}{\ell(\psi,r,R)}\left(1+\frac{H(\phi',\lambda')}{R}+\frac{H^2(\phi',\lambda')}{3R^2}\right)d\sigma', \quad \text{(D.1)}
\end{aligned}$$

with H the height of the topography.

D.1 The exterior potential of the topography

In order to derive the exterior potential of the topography, we use the expansion of the inverse distance (equation 8.7):

$$\frac{1}{\ell} = \sum_{n=0}^{\infty} \frac{1}{r'} \left(\frac{r'}{r}\right)^{n+1} P_n(\cos\psi) = \sum_{n=0}^{\infty} \frac{1}{r} \left(\frac{r'}{r}\right)^{n} P_n(\cos\psi).$$

tasainen suppeneminen [1]

This expansion converges uniformly[1] with respect to ψ if $r > r'$. In the following, we shall assume convergence throughout, dangerous as that may be especially close to a jagged topographic surface. For the philosophically inclined, read Moritz (1980).

Substitution yields

$$\begin{aligned}
V^{\text{ext}}_{\text{top}}(\phi,\lambda,r) &= G\rho \iiint_{\text{top}} \sum_{n=0}^{\infty} \frac{1}{r} \left(\frac{r'}{r}\right)^{n} P_n(\cos\psi)\, d\mathcal{V}' = \\
&= G\rho \iint_{\sigma} \left(\int_{R}^{R+H(\phi',\lambda')} \sum_{n=0}^{\infty} \frac{1}{r} \left(\frac{r'}{r}\right)^{n} (r')^2\, dr' \right) P_n(\cos\psi)\, d\sigma' = \\
&= G\rho \iint_{\sigma} \left[\sum_{n=0}^{\infty} \frac{1}{r^{n+1}} \frac{1}{n+3} (r')^{n+3} \right]_{r'=R}^{R+H} P_n(\cos\psi)\, d\sigma' = \\
&= G\rho \iint_{\sigma} \sum_{n=0}^{\infty} \frac{1}{r^{n+1}} \frac{1}{n+3} \left((R+H)^{n+3} - R^{n+3} \right) P_n(\cos\psi)\, d\sigma'.
\end{aligned}$$

We now use the following Taylor expansion:

---

[1] Uniform convergence means that, given r and r', for every $\epsilon > 0$ there is an $N_{\min}$ for which

$$\left| \frac{1}{\ell} - \frac{1}{r} \sum_{n=0}^{N} \left(\frac{r'}{r}\right)^{n} P_n(\cos\psi) \right| < \epsilon$$

for all $N > N_{\min}$, and *for all values of* ψ. This is a stronger property than mere convergence.

$$
\begin{aligned}
&(R+H)^{n+3} = \\
&= R^{n+3}\left(1+(n+3)\,\frac{H}{R}+\frac{(n+3)\,(n+2)}{2}\,\frac{H^2}{R^2}+\frac{(n+3)\,(n+2)\,(n+1)}{2\cdot 3}\,\frac{H^3}{R^3}+\cdots\right).
\end{aligned}
\tag{D.2}
$$

Substitution yields

$$
\begin{aligned}
&V_{\text{top}}^{\text{ext}}(\phi,\lambda,r) = G\rho R^2\cdot \\
&\cdot\iint_\sigma \sum_{n=0}^{\infty}\left(\frac{R}{r}\right)^{n+1}\left(\frac{H}{R}+\tfrac{1}{2}\,(n+2)\,\frac{H^2}{R^2}+\tfrac{1}{6}\,(n+2)\,(n+1)\,\frac{H^3}{R^3}+\cdots\right)P_n(\cos\psi)\,d\sigma'.
\end{aligned}
\tag{D.3}
$$

This is thus the *exterior potential* of the topography — or, inside the topographic masses, the *harmonic downwards continuation* of the exterior potential, assuming that this is mathematically possible and does not diverge. In mountainous topography, this may be a problem.

D.2 The interior potential of the topography

In the same way we may derive the equation for the interior potential of the topography, the masses between the sea level and terrain surface. For the spatial distance ℓ between those points we use the *interior expansion*, equation 8.7, valid for $r < r'$:

$$
\frac{1}{\ell} = \frac{1}{r}\sum_{n=0}^{\infty}\left(\frac{r}{r'}\right)^{n+1}P_n(\cos\psi).
$$

Substitute:

$$
V_{\text{top}}^{\text{int}}(\phi,\lambda,r) = G\rho\iiint_{\text{top}}\frac{1}{r}\sum_{n=0}^{\infty}\left(\frac{r}{r'}\right)^{n+1}P_n(\cos\psi)\,d\mathcal{V}' =
$$

$$
= G\rho\iint_\sigma \overbrace{\int_R^{R+H(\phi',\lambda')}\frac{1}{r}\sum_{n=0}^{\infty}\left(\frac{r}{r'}\right)^{n+1}(r')^2\,dr'}^{I}\,P_n(\cos\psi)\,d\sigma'.
$$

Here, the height integral I is

$$I = \int_R^{R+H(\phi',\lambda')} \frac{1}{r} \sum_{n=0}^{\infty} \left(\frac{r}{r'}\right)^{n+1} (r')^2 \, dr' =$$

$$= \left[\sum_{\substack{n=0 \\ n\neq 2}}^{\infty} r^n \left(-\frac{(r')^{-(n-2)}}{n-2} \right) + r^2 \ln r' \right]_{r'=R}^{R+H(\phi',\lambda')} =$$

$$= \sum_{\substack{n=0 \\ n\neq 2}}^{\infty} \frac{r^n}{n-2} \left(R^{-(n-2)} - (R+H)^{-(n-2)} \right) + r^2 \ln \frac{R+H}{R},$$

yielding

$$V_{\text{top}}^{\text{int}}(\phi,\lambda,r) =$$

$$= G\rho \iint_\sigma \left(\sum_{\substack{n=0 \\ n\neq 2}}^{\infty} \frac{r^n}{n-2} \left(R^{-(n-2)} - (R+H)^{-(n-2)} \right) + r^2 \ln \frac{R+H}{R} \right) P_n(\cos\psi) \, d\sigma'.$$

For this we use the Taylor expansion

$$(R+H)^{-(n-2)} =$$

$$= R^{-(n-2)} \left(1 - (n-2)\frac{H}{R} + \frac{(n-2)(n-1)}{2}\frac{H^2}{R^2} - \frac{(n-2)(n-1)n}{2\cdot 3}\frac{H^3}{R^3} + \cdots \right).$$

Also, the special case $n = 2$,

$$r^2 \ln \frac{R+H}{R} = r^2 \left(\frac{H}{R} - \frac{1}{2}\frac{H^2}{R^2} + \frac{1}{3}\frac{H^3}{R^3} - \frac{1}{4}\frac{H^4}{R^4} + \cdots \right) =$$

$$= \frac{r^n}{R^{n-2}} \left(\frac{H}{R} - \frac{n-1}{2}\frac{H^2}{R^2} + \frac{(n-1)n}{2\cdot 3}\frac{H^3}{R^3} - \frac{(n-1)n(n+1)}{2\cdot 3\cdot 4}\frac{H^4}{R^4} + \cdots \right),$$

is cleanly included into the following expression obtained by substitution:

$$V_{\text{top}}^{\text{int}}(\phi,\lambda,r) =$$

$$= G\rho \iint_\sigma \sum_{n=0}^{\infty} \frac{r^n}{R^{n-2}} \left(\frac{H}{R} - \frac{1}{2}(n-1)\frac{H^2}{R^2} + \frac{1}{6}(n-1)n\frac{H^3}{R^3} - \cdots \right) P_n(\cos\psi) \, d\sigma'. \quad \text{(D.4)}$$

D.3 The exterior potential of the condensation layer

This is derived by specialising equation D.3 to the case $H \to 0$, but nevertheless $\rho \to \infty$, so that $\kappa = \rho H$ remains finite. in this limit, all terms containing H^2, H^3 and higher powers go to zero. The result is then

$$V^{\text{ext}}_{\text{cond}}(\phi,\lambda,r) = G\rho R^2 \iint_\sigma \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+1} \frac{H}{R} P_n(\cos\psi)\, d\sigma' = \\ = GR \iint_\sigma \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+1} \kappa P_n(\cos\psi)\, d\sigma'.$$

Earlier on we had a more precise formula 6.4 for κ on the surface of a spherical Earth:

$$\kappa = \rho H \left(1 + \frac{H}{R} + \tfrac{1}{3}\frac{H^2}{R^2}\right). \tag{6.4}$$

Substituting this into the previous yields (see also equation D.1):

$$V^{\text{ext}}_{\text{cond}} = G\rho R^2 \iint_\sigma \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+1} \left(\frac{H}{R} + \frac{H^2}{R^2} + \tfrac{1}{3}\frac{H^3}{R^3}\right) P_n(\cos\psi)\, d\sigma'. \tag{D.5}$$

D.4 Total potential of Helmert condensation

This is obtained by subtracting equations D.5 and D.3 from each other. The result — which applies in the *exterior* space[2] — is

$$\delta V^{\text{ext}}_{\text{Helmert}}(\phi,\lambda,r) = V^{\text{ext}}_{\text{cond}}(\phi,\lambda,r) - V^{\text{ext}}_{\text{top}}(\phi,\lambda,r) = -G\rho R^2 \cdot \\ \cdot \iint_\sigma \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+1} \left(\left(\tfrac{1}{2}(n+2) - 1\right) \frac{H^2}{R^2} + \left(\tfrac{1}{6}(n+2)(n+1) - \tfrac{1}{3}\right) \frac{H^3}{R^3} + \cdots \right) \cdot \\ \cdot P_n(\cos\psi)\, d\sigma' = \\ = -G\rho \iint_\sigma \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+1} \left(\tfrac{1}{2} n H^2 + \tfrac{1}{6} n (n+3) \frac{H^3}{R} + \cdots \right) P_n(\cos\psi)\, d\sigma'.$$

[2] Theoretically speaking, the exterior space is the space outside a sphere that encloses all of the Earth's topography, a so-called *Brillouin sphere*. Practice is less restrictive.

asteosuus-yhtälö

Often, we define the *degree constituents* of powers of height H (compare the degree constituent equation 3.9), as follows:

$$H_n^\nu(\phi,\lambda) \stackrel{\text{def}}{=} \frac{2n+1}{4\pi} \iint_\sigma H^\nu(\phi',\lambda')\, P_n(\cos\psi)\, d\sigma', \tag{D.6}$$

with which it holds that

$$H^\nu(\phi,\lambda) = \sum_{n=0}^{\infty} H_n^\nu(\phi,\lambda).$$

Then

$$\delta V^{\text{ext}}_{\text{Helmert}} = \\ = -4\pi G\rho \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+1} \frac{1}{2n+1} \left(\tfrac{1}{2} n H_n^2 + \tfrac{1}{6} n\,(n+3)\,\frac{H_n^3}{R} + \cdots\right).$$

If the topography is constant, all terms vanish for which $n \neq 0$. In the above expansion, in that case the first and second terms also vanish. In this case $n = 0$, the following terms do not even exist: the expansion D.2 is the binomial expansion

$$(R+H)^3 = R^3 + 3R^2H + 3RH^2 + H^3.$$

So

$$\delta V^{\text{ext}}_{\text{Helmert}} = 0$$

as was to be expected according to section 1.4: condensing a spherical shell will not change the exterior field.

D.4.1 The gravity effect of Helmert condensation

Let us calculate the effect of the Helmert condensation potential on *gravity anomalies*:

$$\Delta g^{\text{ext}}_{\text{Helmert}} = -\frac{\partial}{\partial r}\delta V^{\text{ext}}_{\text{Helmert}} - \frac{2}{r}\delta V^{\text{ext}}_{\text{Helmert}} \approx \\ \approx 4\pi G\rho \sum_{n=0}^{\infty} \frac{1}{2n+1}\left(\frac{-(n+1)}{r} + \frac{2}{r}\right)\left(\frac{R}{r}\right)^{n+1}\left(\tfrac{1}{2} n H_n^2 + \tfrac{1}{6} n\,(n+3)\,\frac{H_n^3}{R} + \cdots\right) = \\ = -4\pi G\rho \cdot \frac{1}{r}\sum_{n=0}^{\infty} \frac{n-1}{2n+1}\left(\frac{R}{r}\right)^{n+1}\left(\tfrac{1}{2} n H_n^2 + \tfrac{1}{6} n\,(n+3)\,\frac{H_n^3}{R} + \cdots\right). \tag{D.7}$$

Now, $n = 1$ also gives a zero result, expected as gravity anomalies do not contain any constituents of degree number 1.

Result D.7 is approximate and not to be used on or close to the topography. Note the strong *dependence upon* n: the gravity effect of Helmert condensation is dominated by short wavelengths, the local features of the topography.

D.4.2 The interior potential of Helmert condensation

This quantity is evaluated *on the level of the geoid*. It represents the *indirect effect* of Helmert condensation, the shift of the geoid surface in space caused by the mass shifts. Subtract equations D.5 and D.4 from each other:

epäsuora vaikutus

$$\begin{aligned}
&\delta V^{\text{int}}_{\text{Helmert}}(\phi,\lambda,R) = V^{\text{ext}}_{\text{cond}}(\phi,\lambda,R) - V^{\text{int}}_{\text{top}}(\phi,\lambda,R) = \\
&= G\rho R^2 \iint_\sigma \sum_{n=0}^{\infty} \left(\frac{H}{R} + \frac{H^2}{R^2} + \tfrac{1}{3}\frac{H^3}{R^3} \right) P_n(\cos\psi)\, d\sigma' - \\
&\quad - G\rho R^2 \iint_\sigma \sum_{n=0}^{\infty} \left(\frac{H}{R} - \tfrac{1}{2}(n-1)\frac{H^2}{R^2} + \tfrac{1}{6}(n-1)\,n\frac{H^3}{R^3} - \cdots \right) P_n(\cos\psi)\, d\sigma' = \\
&\quad = G\rho \iint_\sigma \sum_{n=0}^{\infty} \left(\tfrac{1}{2}(n+1)H^2 - \tfrac{1}{6}(n-2)(n+1)\frac{H^3}{R} + \cdots \right) P_n(\cos\psi)\, d\sigma'.
\end{aligned}$$

Using again the definition of the degree constituents of the powers of height H, equation D.6, we obtain

$$\delta V^{\text{int}}_{\text{Helmert}} = 4\pi G\rho \sum_{n=0}^{\infty} \frac{n+1}{2n+1} \left(\tfrac{1}{2}H_n^2 - \tfrac{1}{6}(n-2)\frac{H_n^3}{R} + \cdots \right),$$

from which one obtains with Bruns equation 5.2 the *indirect effect* of Helmert condensation:

$$\begin{aligned}
\delta N_{\text{Helmert}} &= \frac{\delta V^{\text{int}}_{\text{Helmert}}}{\gamma} = \\
&= \frac{4\pi G\rho}{\gamma} \sum_{n=0}^{\infty} \frac{n+1}{2n+1} \left(\tfrac{1}{2}H_n^2 - \tfrac{1}{6}(n-2)\frac{H_n^3}{R} + \cdots \right). \qquad \text{(D.8)}
\end{aligned}$$

The term $n = 0$ yields the indirect effect of a constant terrain $H = \overline{H} = H_0$: using only the first term inside the parentheses yields

$$\delta N_{\text{Helmert, const}} \approx \frac{2\pi G\rho}{\gamma}\overline{H}^2,$$

which *cannot* be neglected.

D.5 The dipole method

As a sanity test, we may describe the effect of Helmert condensation in first approximation as a *dipole-density layer field* μ. The topographic mass, surface density $\kappa = \rho H$, moves downwards by on average $\frac{1}{2}H$.
3 The effect would be the same if the mean sea level[3] were covered by a double mass-density layer

$$\mu = \tfrac{1}{2}\rho H^2. \tag{D.9}$$

The potential of this layer is, in spherical approximation (equation 1.18):

$$V = G\iint_S \mu\frac{\partial}{\partial n}\left(\frac{1}{\ell}\right)dS \approx GR^2\iint_\sigma \mu\frac{\partial}{\partial n}\left(\frac{1}{\ell}\right)d\sigma.$$

Written more explicitly in spherical geometry:

$$V_P = GR^2\iint_\sigma \mu_Q\frac{\partial}{\partial r_Q}\left(\frac{1}{\ell_{PQ}}\right)d\sigma_Q.$$

We use the expansion into Legendre polynomials, equation 8.7:

$$\frac{1}{\ell_{PQ}} = \frac{1}{r_Q}\sum_{n=0}^{\infty}\left(\frac{r_Q}{r_P}\right)^{n+1}P_n(\cos\psi_{PQ}),$$

differentiate with respect to r_Q, and substitute:

$$V_P = GR^2\iint_\sigma \frac{1}{r_Q^2}\mu_Q\sum_{n=0}^{\infty} n\left(\frac{r_Q}{r_P}\right)^{n+1}P_n(\cos\psi_{PQ})\,d\sigma_Q.$$

[3] In fact, a better place for this replacement layer would be the $\frac{1}{4}H$ level.

By substituting into this equation D.9 for the double mass-density layer μ_Q we obtain, by taking the limit $r_P, r_Q \downarrow R$:

$$V = \frac{1}{4\pi} \sum_{n=0}^{\infty} n \iint_\sigma (2\pi G\rho H)\, H P_n(\cos\psi)\, d\sigma' = \\ = \frac{1}{4\pi} \sum_{n=0}^{\infty} n \iint_\sigma A_B H P_n(\cos\psi)\, d\sigma'.$$

Here, we have left off the designations P and Q again as they are no longer needed for clarity.

The symbol A_B denotes the attraction of a Bouguer plate of thickness H and matter density ρ.

Let us develop the quantity $(A_B H)$ into a spherical-harmonic expansion. According to degree constituent equation 3.9: pallofunktiokehitelmä

$$(A_B H)_n = \frac{2n+1}{4\pi} \iint_\sigma (A_B H)\, P_n(\cos\psi)\, d\sigma',$$

yielding

$$V = \sum_{n=0}^{\infty} \frac{n}{2n+1} (A_B H)_n \approx \tfrac{1}{2} (A_B H),$$

at least for the higher n values: regionally though not globally.

Thus we obtain again an estimate for the indirect effect of Helmert condensation. In geoid computation by means of this method this represents the shift in geoid surface caused by the condensation, which must be undone, that is, accounted for with the opposite algebraic sign. In other words, when looked upon as a *remove–restore* method, it constitutes its "restore" step: poistamis-entistämismenetelmä

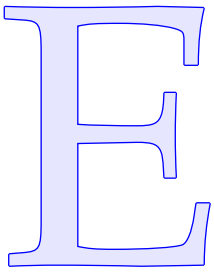
$$\delta N_{\text{Helmert}} = \frac{V}{\gamma} \approx \tfrac{1}{2} \frac{A_B H}{\gamma} = \frac{\pi G\rho H^2}{\gamma}.$$

For comparison, the more precise expansion D.8 yields in approximation for larger n values

$$\delta N_{\text{Helmert}} \approx \frac{4\pi G\rho}{\gamma} \cdot \tfrac{1}{2} \sum_{n=0}^{\infty} \frac{n+1}{2n+1} H_n^2 \approx \frac{\pi G\rho}{\gamma} \sum_{n=0}^{\infty} H_n^2 = \frac{\pi G\rho H^2}{\gamma},$$

essentially the same result.

The Laplace equation in spherical co-ordinates



E.1 Derivation

Consider a small volume element with sizes in co-ordinate directions of $\Delta\phi$, $\Delta\lambda$, and Δr. Look at the difference in flux of vector field $\mathbf{a} \stackrel{\text{def}}{=} \nabla V$ between what comes in and what goes out through opposite faces. vuo

We do the analogue of what was shown in subsection 1.12.4, using a body or volume element with surfaces aligned along co-ordinate lines, allowing the size of the element to go to zero in the limit, and exploiting the divergence theorem 1.19 of Gauss. The quantity $\operatorname{div} \mathbf{a} = \Delta V$ is a *source density* in space, and its average value multiplied by the volume of an element must equal the total flux through the surfaces of the element.

Define at the location of the body an orthonormal basis $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ of type "north-east-up". The vector $\boldsymbol{e}_1$ points to the local north, the vector $\boldsymbol{e}_2$ to the east, and the vector $\boldsymbol{e}_3$ "up", in the radial direction. We may write ortonormaali kanta

$$\mathbf{a} = a_1\boldsymbol{e}_1 + a_2\boldsymbol{e}_2 + a_3\boldsymbol{e}_3.$$

Part of the difference in flux f between opposing faces is due to a change in the normal component of $\mathbf{a}$ between the faces, part is due to a difference in face surface area ω:

$$f^+ - f^- \approx \overbrace{\omega\,(a^+ - a^-)}^{\text{I}} + \overbrace{a\,(\omega^+ - \omega^-)}^{\text{II}}.$$

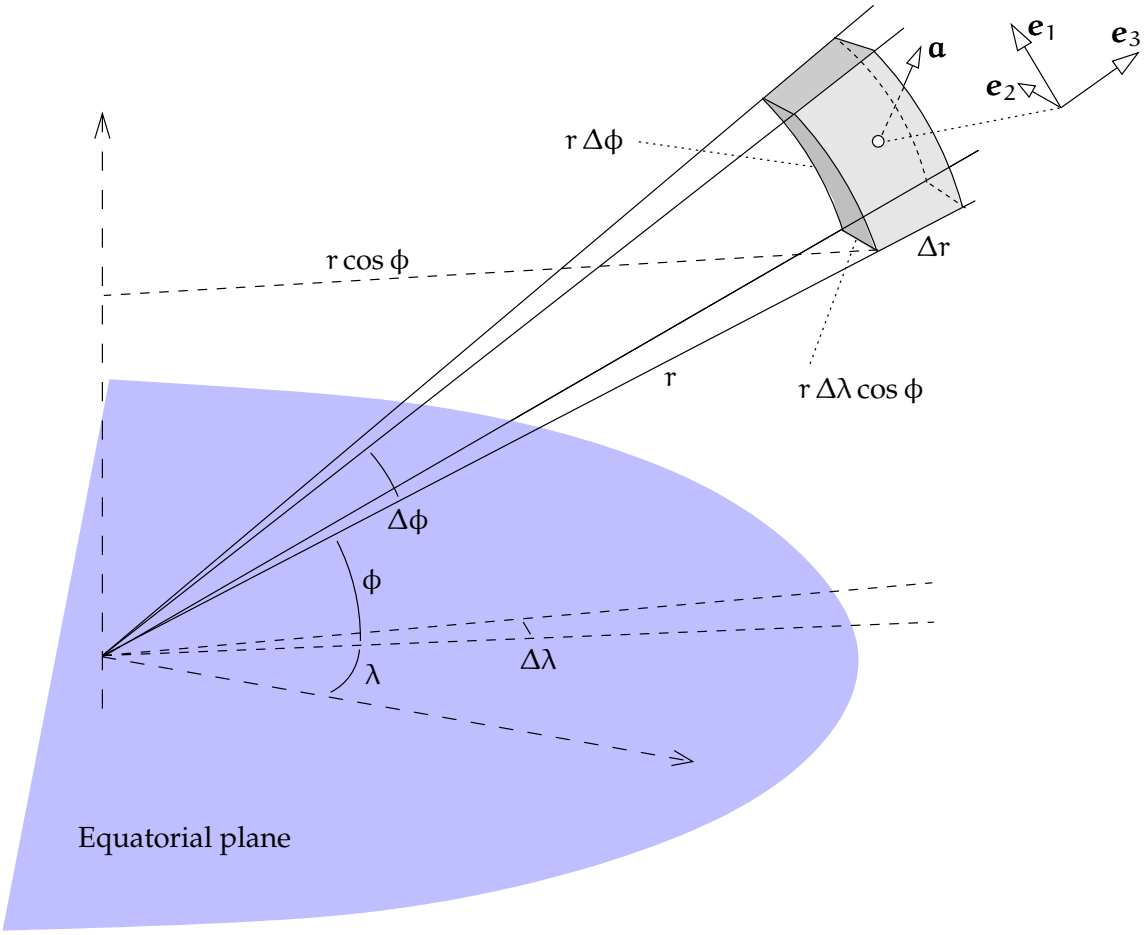


FIGURE E.1. The Gauss divergence theorem applied to a co-ordinate aligned volume element.

See figure E.1.

- Latitudinal direction, ϕ, "south–north":

$$\omega_\phi^- = r\cos\phi\,\Delta r\,\Delta\lambda, \qquad \omega_\phi^+ = r\cos(\phi+\Delta\phi)\,\Delta r\,\Delta\lambda,$$

difference

$$\omega_\phi^+ - \omega_\phi^- \approx -r\sin\phi\,\Delta\phi\cdot\Delta r\,\Delta\lambda.$$

Multiply by

$$a_1 = \frac{\partial V}{\partial(r\phi)} = \frac{1}{r}\frac{\partial V}{\partial\phi}$$

and divide by element volume $r^2 \cos\phi \, \Delta r \, \Delta\phi \, \Delta\lambda$, yielding

$$\Delta^{\mathrm{II}}_{\phi} V = -\frac{\tan\phi}{r^2} \frac{\partial V}{\partial \phi}.$$

This of course in addition to the first contribution

$$\Delta^{\mathrm{I}}_{\phi} V = \langle \nabla a_1 \cdot \boldsymbol{e}_1 \rangle = \frac{a_1^+ - a_1^-}{r \cdot \Delta\phi},$$

with

$$a_1^+ - a_1^- = \left[\frac{\partial V}{\partial (r\phi)} \right]_-^+ = \frac{1}{r} \left[\frac{\partial V}{\partial \phi} \right]_-^+,$$

yielding

$$\Delta^{\mathrm{I}}_{\phi} V = \frac{1}{r} \cdot \frac{1}{r} \cdot \frac{\left[\frac{\partial}{\partial \phi} V \right]_-^+}{\Delta\phi} \approx \frac{1}{r^2} \frac{\partial^2 V}{\partial \phi^2}.$$

- Longitudinal direction, λ, "west–east": no change in surface area $\omega_\lambda = r \, \Delta r \, \Delta\phi$ because of rotational symmetry:

$$\Delta^{\mathrm{II}}_{\lambda} V = 0.$$

We only have

$$\Delta^{\mathrm{I}}_{\lambda} V = \langle \nabla a_2 \cdot \boldsymbol{e}_2 \rangle = \frac{a_2^+ - a_2^-}{r \cos\phi \cdot \Delta\lambda},$$

with

$$a_2^+ - a_2^- = \left[\frac{\partial V}{\partial (\lambda r \cos\phi)} \right]_-^+ = \frac{1}{r \cos\phi} \left[\frac{\partial V}{\partial \lambda} \right]_-^+.$$

Substitution yields

$$\Delta^{\mathrm{I}}_{\lambda} V = \frac{1}{r \cos\phi} \cdot \frac{1}{r \cos\phi} \cdot \frac{\left[\frac{\partial}{\partial \lambda} V \right]_-^+}{\Delta\lambda} \approx \frac{1}{r^2 \cos^2\phi} \frac{\partial^2 V}{\partial \lambda^2}.$$

- In the radial direction, the surface areas of opposing faces — "inner–outer" — are

$$\omega_r^- = r^2 \cos\phi \, \Delta\phi \, \Delta\lambda, \qquad \omega_r^+ = (r + \Delta r)^2 \cos\phi \, \Delta\phi \, \Delta\lambda,$$

the difference being

$$\omega_r^+ - \omega_r^- \approx 2r\,\Delta r \cdot \cos\phi\,\Delta\phi\,\Delta\lambda.$$

Multiply by

$$a_3 = \frac{\partial V}{\partial r}$$

and divide by the volume of the element $r^2 \cos\phi\,\Delta r\,\Delta\phi\,\Delta\lambda$, yielding for the second contribution to the Laplace operator

$$\Delta_r^{II} V = \frac{2}{r}\frac{\partial V}{\partial r}.$$

This in addition to the first contribution

$$\Delta_r^I V = \langle \nabla a_3 \cdot \boldsymbol{e}_3 \rangle = \frac{a_3^+ - a_3^-}{\Delta r} = \frac{\left[\frac{\partial}{\partial r} V\right]_-^+}{\Delta r} \approx \frac{\partial^2 V}{\partial r^2}.$$

All of this gives us the end result

$$\begin{aligned} \Delta V = \Delta_r^I V + \Delta_\lambda^I V + \Delta_\phi^I V + \Delta_r^{II} V + \Delta_\phi^{II} V = \\ = \frac{\partial^2 V}{\partial r^2} + \frac{1}{r^2\cos^2\phi}\frac{\partial^2 V}{\partial \lambda^2} + \frac{1}{r^2}\frac{\partial^2 V}{\partial \phi^2} + \frac{2}{r}\frac{\partial V}{\partial r} - \frac{\tan\phi}{r^2}\frac{\partial V}{\partial \phi}, \end{aligned} \quad \text{(E.1)}$$

equivalent to equation 2.9.

E.2 Solution

E.2.1 Separating the radial dependency

Let us attempt separation of variables as follows:

$$V(\phi, \lambda, r) = R(r)\,Y(\phi, \lambda).$$

Substitution into equation E.1 and multiplication by r^2/RY yields

$$\frac{1}{R}\left(r^2\frac{\partial^2 R}{\partial r^2} + 2r\frac{\partial R}{\partial r}\right) = -\frac{1}{Y}\left(\frac{1}{\cos^2\phi}\frac{\partial^2 Y}{\partial \lambda^2} + \frac{\partial^2 Y}{\partial \phi^2} - \tan\phi\frac{\partial Y}{\partial \phi}\right).$$

This must again apply for all values r and ϕ and thus both expressions can only be equal to a constant, p. This yields two equations:

$$\left(r^2\frac{\partial^2 R}{\partial r^2}+2r\frac{\partial R}{\partial r}\right)-pR=0,$$
$$\left(\frac{1}{\cos^2\phi}\frac{\partial^2 Y}{\partial\lambda^2}+\frac{\partial^2 Y}{\partial\phi^2}-\tan\phi\frac{\partial Y}{\partial\phi}\right)+pY=0.$$

For the first equation we try a power law,

$$R(r)=r^q,$$

yielding

$$q\,(q-1)\,r^q+2qr^q-pr^q=0 \quad\Longrightarrow\quad (q\,(q+1)-p)\,r^q=0$$

with the solution

$$p=q\,(q+1)\,.$$

Solving the second equation for $Y(\phi,\lambda)$,

$$\left(\frac{1}{\cos^2\phi}\frac{\partial^2 Y}{\partial\lambda^2}+\frac{\partial^2 Y}{\partial\phi^2}-\tan\phi\frac{\partial Y}{\partial\phi}\right)+q\,(q+1)\,Y=0, \qquad \text{(E.2)}$$

is trickier. It turns out that q must be an integer. One finds, for $n\in\mathbb{N}_0$, that there are non-negative solutions $q=n$ and negative solutions $q=-(n+1)$, with $n=0,1,2,\ldots$. With this, the full set of special solutions is

$$V_n^{\text{int}}(\phi,\lambda,r)=r^nY_n(\phi,\lambda),\quad V_n^{\text{ext}}(\phi,\lambda,r)=\frac{Y_n(\phi,\lambda)}{r^{n+1}},\qquad n\in\mathbb{N}_0,$$

equations 2.10.

E.2.2 Solving for surface harmonics

Both solutions q, the non-negative and the negative one, yield on substitution into equation E.2 the same equation for n:

$$\left(\frac{1}{\cos^2\phi}\frac{\partial^2 Y}{\partial\lambda^2}+\frac{\partial^2 Y}{\partial\phi^2}-\tan\phi\frac{\partial Y}{\partial\phi}\right)+n\,(n+1)\,Y=0.$$

We attempt separation of variables:

$$Y(\phi, \lambda) = F(\phi)\, L(\lambda).$$

Substitution and multiplication by $\cos^2\phi \big/ FL$ yields

$$\frac{\cos^2\phi}{F}\left(\frac{\partial^2 F}{\partial\phi^2} - \tan\phi\frac{\partial F}{\partial\phi} + n\,(n+1)\,F\right) = -\frac{1}{L}\frac{\partial^2 L}{\partial\lambda^2}.$$

Both sides must be again equal to the same constant, which we shall assume to be positive and call m^2:

$$\frac{\partial^2 F}{\partial\phi^2} - \tan\phi\frac{\partial F}{\partial\phi} + \left(n\,(n+1) - \frac{m^2}{\cos^2\phi}\right)F = 0, \quad \frac{\partial^2 L}{\partial\lambda^2} + m^2 L = 0.$$

The first equation is known as Legendre's equation. Its solutions are the Legendre functions $P_{nm}(\sin\phi)$, with the integer $m = 0, 1, \ldots, n$.

harmoninen värähtelijä [1]

The second is the classical *harmonic oscillator*, with solutions[1]

$$L_{m,1}(\lambda) = \cos m\lambda, \qquad L_{m,2}(\lambda) = \sin m\lambda.$$

pinta-pallofunktio

With this, we find for the surface spherical harmonics the linear combinations

$$Y_n(\phi, \lambda) = \sum_{m=0}^{n} P_{nm}(\sin\phi)\,(a_{nm}\cos m\lambda + b_{nm}\sin m\lambda)\,.$$

The general solution is now formed as follows:

$$V^{\text{int}}(\phi, \lambda, r) = \sum_{n=0}^{\infty} r^n \sum_{m=0}^{n} P_{nm}(\sin\phi)\,(a_{nm}\cos m\lambda + b_{nm}\sin m\lambda)\,,$$

$$V^{\text{ext}}(\phi, \lambda, r) = \sum_{n=0}^{\infty} \frac{1}{r^{n+1}} \sum_{m=0}^{n} P_{nm}(\sin\phi)\,(a_{nm}\cos m\lambda + b_{nm}\sin m\lambda)\,.$$

pallofunktio-kerroin

Here, a_{nm} and b_{nm} are the spherical-harmonic coefficients specifying the linear combination of special solutions. Only the second solution is physically realistic for representing the Earth's exterior gravitational field, going to zero at infinity $r \to \infty$.

[1] This also explains why m must be an integer: the longitude λ is circular with a period of 2π.

Bibliography

A

M. Abrehdary, L. E. Sjöberg, and M. Bagherbandi. The spherical terrain correction and its effect on the gravimetric-isostatic Moho determination. *Geophysical Journal International*, 204(1):262–273, 2016. URL https://doi.org/10.1093/gji/ggv450. 141

G. d'Agostino, S. Desogus, A. Germak, C. Origlia, D. Quagliotti, G. Berrino, G. Corrado, V. d'Errico, and G. Ricciardi. The new IMGC-02 transportable absolute gravimeter: Measurement apparatus and applications in geophysics and volcanology. *Annals of Geophysics*, 51(1):39–49, 2008. URL https://doi.org/10.4401/ag-3038. 312

Altimetry, Retracking. Radar altimetry tutorial & toolbox. ESA and CNES. URL http://www.altimetry.info/radar-altimetry-tutorial/data-flow/data-processing/retracking/. Accessed 1st March, 2022. 373

O. B. Andersen, P. Knudsen, and P. Berry. The DNSC08GRA global marine gravity field from double retracked satellite altimetry. *Journal of Geodesy*, 84: 191–199, 2010. URL https://doi.org/10.1007/s00190-009-0355-9. 249

M. de Angelis, A. Bertoldi, L. Cacciapuoti, A. Giorgini, G. Lamporesi, M. Prevedelli, G. Saccorotti, F. Sorrentino, and G. M. Tino. Precision gravimetry with atomic sensors. *Measurement Science and Technology*, 20(2): 022001, 2009. URL http://dx.doi.org/10.1088/0957-0233/20/2/022001. 312

B

G. Balmino, N. Vales, S. Bonvalot, and A. Briais. Spherical harmonic modeling to ultra-high degree of Bouguer and isostatic anomalies. *Journal of Geodesy*, 86:499–520, 2012. URL https://doi.org/10.1007/s00190-011-0533-4. 64, 140

T. B. Bedada. *Absolute geopotential height system for Ethiopia*. PhD thesis, University of Edinburgh, 2010. URL https://era.ed.ac.uk/bitstream/handle/1842/4726/Bedada2010_small.pdf. Accessed 22nd September, 2021. 249

N. Benitez, T. J. Broadhurst, H. C. Ford, M. Clampin, G. Hartig, G. D. Illingworth, et al. Hubble Looks Through Cosmic Zoom Lens, 2003. URL https://esahubble.org/images/op00301a/. © 2003 ESA/Hubble (CC BY 4.0). Accessed 22nd September, 2021. 2

BGI, EGM2008. EGM2008 anomaly maps visualization. URL http://bgi.obs-mip.fr/data-products/outils/egm2008-anomaly-maps-visualization/. Accessed 17th September, 2021. 126, 136

BGI, WGM2012. WGM2012 maps visualization/extraction. URL http://bgi.obs-mip.fr/data-products/outils/wgm2012-maps-visualizationextraction/. Accessed 17th September, 2021. 154

M. Bilker-Koivula. Development of the Finnish Height Conversion Surface FIN2005N00. *Nordic Journal of Surveying and Real Estate Research*, 7(1):76–88, 2010. URL https://journal.fi/njs/article/download/3663/3432. Accessed 22nd September, 2021. 403

M. Bilker-Koivula and M. Ollikainen. Suomen geoidimallit ja niiden käyttäminen korkeuden muunnoksissa. Research note (in Finnish) 29, Finnish Geodetic Institute, 2009. URL https://www.maanmittauslaitos.fi/sites/maanmittauslaitos.fi/files/fgi/GLtiedote29.pdf. Accessed 11th May, 2019. 247, 332

G. P. Bottoni and R. Barzaghi. Fast collocation. *Bulletin Géodésique*, 67(2): 119–126, 1993. URL https://doi.org/10.1007/BF01371375. 291

V. V. Brovar, M. I. Yurkina, M. Heifets, M. S. Molodensky, and H. Moritz. M. S. Molodensky In Memoriam. Online PDF, Mitteilungen der geodätischen Institute der Technischen Universität Graz Folge 88, 2000. URL

http://www.helmut-moritz.at/SciencePage/Molodensky.pdf. Helmut Moritz and Maria I. Yurkina, editors. Accessed 1st March, 2020. 170

J. M. Brozena. The Greenland Aerogeophysics Project: Airborne gravity, topographic and magnetic mapping of an entire continent. International Association of Geodesy Symposia 110, pages 203–214, Vienna, Austria, 20th August, 1992. Springer, New York, NY. URL https://doi.org/10.1007/978-1-4613-9255-2_19. 321

J. M. Brozena and M. F. Peters. State-of-the-art airborne gravimetry. International Association of Geodesy Symposia 113, pages 187–197, Graz, Austria, 1994. Springer-Verlag. URL https://doi.org/10.1007/978-3-642-79721-7_20. 321

J. M. Brozena, M. F. Peters, and R. Salman. Arctic airborne gravity measurements program. In Segawa et al. (1996), pages 131–146. URL https://doi.org/10.1007/978-3-662-03482-8_20. 321

H. Bruns. *Die Figur der Erde: Ein Beitrag zur europäischen Gradmessung*. Stankievicz, Berlin, 1878. URL https://play.google.com/books/reader?id=DPo-AAAAYAAJ&hl=en&pg=GBS.PP5. Accessed 31st January, 2020. 94

B. Bucha, C. Hirt, and M. Kuhn. Cap integration in spectral gravity forward modelling: near- and far-zone gravity effects via Molodensky's truncation coefficients. *Journal of Geodesy*, 93:65–83, 2019. URL https://doi.org/10.1007/s00190-018-1139-x. 223

C

L. Caesar, S. Rahmstorf, A. Robinson, G. Feulner, and V. S. Saba. Observed fingerprint of a weakening Atlantic Ocean overturning circulation. *Nature*, 556:191–196, 2018. URL https://doi.org/10.1038/s41586-018-0006-5. 339

H. Cavendish. Experiments to determine the density of the earth. *Philosophical Transactions of the Royal Society*, 88, 1798. URL https://doi.org/10.1098/rstl.1798.0022. 4

CHAMP Mission. CHAMP – CHAllenging Minisatellite Payload. Deutsches Geoforschungszentrum, Helmholtz-Zentrum Potsdam. URL https://www.gfz-potsdam.de/champ/. Accessed 1st March, 2020. 377

Climate Research Unit. University of East Anglia. URL https://crudata.uea.ac.uk/cru/data/soi/. Accessed 13th January, 2022. 354

B. J. Coakley, S. C. Kenyon, and R. Forsberg. Updating the Arctic Gravity Project grid with new airborne and Extended Continental Shelf data. *AGU Fall Meeting Abstracts*, page C3, December 2013. URL https://ui.adsabs.harvard.edu/abs/2013AGUFM.G13C..03C/abstract. Accessed 17th January, 2020. 321

J. J. O'Connor and E. F. Robertson. George Green (1793–1841). MacTutor History of Mathematics archive, School of Mathematics and Statistics, University of St Andrews, Scotland, 1998. URL https://mathshistory.st-andrews.ac.uk/Biographies/Green/. Accessed 22nd September, 2021. 29

D

B. D. DeJong, P. R. Bierman, W. L. Newell, Tammy, M. Rittenour, S. A. Mahan, G. Balco, and D. H. Rood. Pleistocene relative sea levels in the Chesapeake Bay region and their implications for the next century. *GSA Today*, 25(8): 4–10, 2015. URL https://doi.org/10.1130/GSATG223A.1. 342

H. Denker. Evaluation and improvement of the EGG97 quasigeoid model for Europe by GPS and leveling data. In Vermeer and Ádám (1998), pages 53–61. 402

E

Eiffel Tower, 72 names. List of the 72 names on the Eiffel Tower. URL https://en.wikipedia.org/wiki/List_of_the_72_names_on_the_Eiffel_Tower. Accessed 7th April, 2019. 16, 17, 34, 46, 57, 90, 433

M. Ekman. Postglacial rebound and sea level phenomena with special reference to Fennoscandia and the Baltic Sea. In Kakkuri (1993), pages 7–70. 337, 394

Encyclopaedia Britannica, Moho. URL https://www.britannica.com/science/Moho. Accessed 22nd January, 2020. 159

L. Eötvös. *Three Fundamental Papers of Loránd Eötvös*. Loránd Eötvös Geophysical Institute of Hungary, 1998. ISBN 963-7135-02-2. Editor Zoltán Szabó. 323

European geoid calculations. Leibniz Universität Hannover, Institute of Geodesy. URL https://www.ife.uni-hannover.de/en/research/main-research-focus/regional-gravity-field-and-geoid-modelling/european-geoid-calculations/. Accessed 11th May, 2019. 402

F

G. Farmelo. *The Strangest Man*. Basic Books, reprint edition, 2011. ISBN 978-0-4650-2210-6. 27

W. E. Farrell and J. A. Clark. On postglacial sea level. *Geophysical Journal of the Royal Astronomical Society*, 46(3):647–667, 1976. URL https://doi.org/10.1111/j.1365-246X.1976.tb01252.x. 342

W. E. Featherstone. Software for computing five existing types of deterministically modified integration kernel for gravimetric geoid determination. *Computers and Geosciences*, 29:183–193, 2003. URL https://doi.org/10.1016/S0098-3004(02)00074-2. 225

FFTW Home Page. URL https://www.fftw.org. Accessed 15th May, 2019. 441

R. Forsberg. A study of terrain reductions, density anomalies and geophysical inversion methods in gravity field modelling. Report 355, Ohio State University, Department of Geodetic Science and Surveying, 1984. URL https://earthsciences.osu.edu/sites/earthsciences.osu.edu/files/report-355.pdf. Accessed 17th February, 2020. 251

R. Forsberg and J. Kaminskis. Geoid of the Nordic and Baltic region from gravimetry and satellite altimetry. In Segawa et al. (1996), pages 540–547. URL https://doi.org/10.1007/978-3-662-03482-8_72. 403

R. Forsberg and G. Strykowski. NKG Gravity Data Base and NKG Geoid. Technical University of Denmark, DTU Space, National Space Institute, 2010. URL https://www.nordicgeodeticcommission.com/wp-content/uploads/20

14/10/8-WG_geoid_2010_March_presentation_ForsbergStrykowski.pdf. Accessed 11th May, 2019. 403

R. Forsberg, A. V. Olesen, F. Ferraccioli, T. Jordan, H. Corr, and K. Matsuoka. PolarGap 2015/16 - Filling the GOCE polar gap in Antarctica and ASIRAS flight around South Pole. Final report, European Space Agency, 2017. URL https://earth.esa.int/documents/10174/134665/PolarGap-2015-2016-final-report. Accessed 8th January, 2020. 380

R. Forsberg and C. C. Tscherning. An overview manual for the GRAVSOFT Geodetic Gravity Field Modelling Programs, 2008. URL https://www.academia.edu/9206363/An_overview_manual_for_the_GRAVSOFT_Geodetic_Gravity_Field_Modelling_Programs. Accessed 11th May, 2019. 247

R. Forsberg and M. Vermeer. A generalized Strang van Hees approach to fast geopotential inversion. *Manuscripta geodaetica*, 17:302–314, 1992. 243

J.-P. Friedelmeyer. Du côté des lettres (2) : une lettre de Sophie Germain à Carl Friedrich Gauss (20 février 1807), et la réponse de celui-ci (30 avril 1807), 2014. URL https://images.math.cnrs.fr/Du-cote-des-lettres-une-lettre-de-Sophie-Germain-a-Carl-Friedrich-Gauss-20. Accessed 22nd September, 2021. 94

M. Frigo and S. G. Johnson. The design and implementation of FFTW3. *Proceedings IEEE*, 93(2):216–231, 2005. URL http://www.fftw.org/fftw-paper-ieee.pdf. Accessed 14th February, 2020. 441

G

E. S. Garcia, D. T. Sandwell, and W. H. Smith. Retracking CryoSat-2, Envisat and Jason-1 radar altimetry waveforms for improved gravity field recovery. *Geophysical Journal International*, 2014. URL https://doi.org/10.1093/gji/ggt469. 249

A. Gatti, M. Reguzzoni, F. Sansò, and F. Migliaccio. Space-wise grids of gravity gradients from GOCE data at nominal satellite altitude. Presented at the 5th International GOCE User Workshop, UNESCO, Paris, France, 25th – 28th November, 2014. URL

https://www.researchgate.net/publication/275029640_SPACE-WISE_GRIDS_OF_GRAVITY_GRADIENTS_FROM_GOCE_DATA_AT_NOMINAL_SATELLITE_ALTITUDE. Accessed 11th May, 2019. 288

O. R. Godø, A. Samuelsen, G. J. Macaulay, R. Patel, S. S. Hjøllo, J. Horne, S. Kaartvedt, and J. A. Johannessen. Mesoscale eddies are oases for higher trophic marine life. *PLoS One*, 7(1):e30161, 2012. URL https://doi.org/10.1371/journal.pone.0030161. 340

GRACE Follow-On Mission. Jet Propulsion Laboratory. URL https://gracefo.jpl.nasa.gov/mission/overview/. Accessed 1st March, 2020. 379

GRACE Mission. Measuring Earth's Surface Mass and Water Changes. Jet Propulsion Laboratory. URL https://grace.jpl.nasa.gov/. Accessed 1st March, 2020. 378

GRACE Mission, hydrology. NASA. URL https://commons.wikimedia.org/wiki/File:Global_Gravity_Anomaly_Animation_over_LAND.gif. Accessed 1st March, 2020. 379

G. Green. *An Essay on the Application of Mathematical Analysis to the Theories of Electricity and Magnetism*. 1828. URL https://play.google.com/books/reader?id=GwYXAAAAYAAJ. Accessed 23rd April, 2019. 29

Green's Windmill. Green's windmill and science centre. URL https://www.greensmill.org.uk/. Accessed 10th May, 2019. 29

GWR Instruments, Inc., *i*GRAV® Gravity Sensors. URL https://www.gwrinstruments.com/igrav-gravity-sensors.html. Accessed 22nd September, 2022. 317

H

R. Haagmans, E. de Min, and M. van Gelderen. Fast evaluation of convolution integrals on the sphere using 1D FFT, and a comparison with existing methods for Stokes' integral. *Manuscripta geodaetica*, 18:227–241, 1993. 243

P. Häkli, J. Puupponen, H. Koivula, and M. Poutanen. Suomen geoidimallit ja niiden käyttäminen korkeuden muunnoksissa. Research note (in Finnish) 30, Finnish Geodetic Institute, 2009. URL https://www.maanmittauslaitos.fi/sites/maanmittauslaitos.fi/files/fgi/GLtiedote30.pdf. Accessed 26th January, 2020. 332

J. C. Harrison and M. Dickinson. Fourier transform methods in local gravity modelling. *Bulletin Géodésique*, 63:149–166, 1989. URL https://doi.org/10.1007/BF02519148. 251

M. Heikkinen. Solving the shape of the Earth by using digital density models. Report 81:2, Finnish Geodetic Institute, Helsinki, 1981. 101, 102, 104, 318

W. A. Heiskanen. The latest achievements of physical geodesy. *Journal of Geophysical Research*, 65(9):2827–2836, 1960. URL https://doi.org/10.1029/JZ065i009p02827. 155

W. A. Heiskanen and H. Moritz. *Physical Geodesy*. W. H. Freeman and Company, San Francisco, London, 1967. 35, 53, 57, 68, 71, 79, 80, 81, 94, 95, 97, 98, 104, 105, 106, 114, 118, 153, 155, 157, 177, 179, 180, 184, 193, 194, 201, 202, 206, 211, 293, 404, 435

W. F. Hermans. *Beyond Sleep*. Harry N. Abrams, reprint edition, 2007. ISBN 978-1-5856-7583-8. 146

C. Hirt and M. Kuhn. Band-limited topographic mass distribution generates full-spectrum gravity field: Gravity forward modeling in the spectral and spatial domains revisited. *Journal of Geophysical Research: Solid Earth*, 119, 2014. URL https://doi.org/10.1002/2013JB010900. 140

R. A. Hirvonen. *The continental undulations of the geoid*. PhD thesis, Helsinki University of Technology, 1934. Finnish Geodetic Institute publication 19. 403

R. A. Hirvonen. *Relative Bestimmungen der Schwerkraft in Finnland in den Jahren 1931, 1933 und 1935*. Publication 23, Finnish Geodetic Institute, Helsinki, 1937. 300

B. Hofmann-Wellenhof and H. Moritz. *Physical Geodesy*. Springer-Verlag Wien GmbH, 2006. Second, revised edition. 404

Humboldt University Berlin. Friedrich Robert Helmert with a relative pendulum, 2017. URL https://www.researchgate.net/publication/318994932_Friedrich_Robert_Helmert_founder_of_modern_geodesy_on_the_occasion_of_the_centenary_of_his_death. © 2017 Humboldt-Universität zu Berlin, Universitätsbibliothek (CC BY 3.0). Accessed 19th May, 2019. 144

E. Hytönen. *Absolute gravity measurement with long wire pendulum*. PhD thesis, University of Helsinki, 1972. Finnish Geodetic Institute publication 75. 301

I

International Intercomparison of Absolute Gravimeters. University of Luxembourg, European Center for Geodynamics and Seismology. URL http://www.ecgs.lu/international-intercomparison-of-absolute-gravimeters/. Accessed 6th February, 2020. 314

ISG, Geoid Schools. International Service for the Geoid. URL http://www.isgeoid.polimi.it/Schools/schools.html. Accessed 1st March, 2020. 247

K

E. Kääriäinen. *The Second Levelling of Finland in 1935–1955*. Publication 61, Finnish Geodetic Institute, Helsinki, 1966. 165, 166, 180

J. Kääriäinen and H. Ruotsalainen. *Tilt Measurements in the Underground Laboratory Lohja 2, Finland, in 1977–1987*. Publication 110, Finnish Geodetic Institute, Helsinki, 1989. 394

J. Kakkuri, editor. *Geodesy and Geophysics, lecture notes, NKG Autumn School 1992*, Finnish Geodetic Institute publication 115, 1993. 462, 473, 474

J. Kakkuri. *Surveyor of the Globe – Story of the Life of V. A. Heiskanen*. Finnish National Land Survey, 2008. URL https://readymag.com/u95015526/508134/. Accessed 13th May, 2019. 146, 404

S. C. Kenyon, R. Forsberg, A. V. Olesen, and S. A. Holmes. NGA's use of aerogravity to advance the next generation of Earth Gravitational Models.

AGU Fall Meeting Abstracts, page A4, December 2012. URL https://ui.adsabs.harvard.edu/abs/2012AGUFM.G12A..04K/abstract. Accessed 11th May, 2019. 321

M. G. Kogan, M. Diament, A. Bulot, and G. Balmino. Thermal isostasy in the South Atlantic Ocean from geoid anomalies. *Earth and Planetary Science Letters*, 74:280–290, 1985. URL https://doi.org/10.1016/0012-821X(85)90028-7. 155

M. Kuhn, W. Featherstone, and J. Kirby. Complete spherical Bouguer gravity anomalies over Australia. *Australian Journal of Earth Sciences*, 56(2):213–223, 2009. URL https://espace.curtin.edu.au/handle/20.500.11937/34751. Accessed 1st March, 2020. 140, 141

A. Kuivamäki, P. Vuorela, and M. Paananen. Indications of postglacial and recent bedrock movements in Finland and Russian Karelia. *Report* YST-99, Geological Survey of Finland, 1998. URL http://tupa.gtk.fi/julkaisu/ydinjate/yst_099.pdf. Accessed 17th January, 2020. 347

L

A. E. H. Love. The Yielding of the Earth to Disturbing Forces. *Proceedings of the Royal Society of London A*, 82(551):73–88, 1909. URL https://doi.org/10.1098/rspa.1909.0008. 391

B. Lu, F. Barthelmes, S. Petrovic, C. Förste, F. Flechtner, Z. Luo, K. He, and M. Li. Airborne gravimetry of GEOHALO mission: Data processing and gravity field modeling. *Journal of Geophysical Research: Solid Earth*, 122: 10 586–10 604, 2017. URL https://doi.org/10.1002/2017JB014425. 321

M

J. Mäkinen, A. Engfeldt, L. Engman, B. G. Harsson, T. Oja, S. Rekkedal, K. Røthing, P. Rouhiainen, H. Ruotsalainen, H. Skatt, G. Strykowski, H. Virtanen, K. Wieczerkowski, and D. Wolf. The Fennoscandian Land Uplift Gravity Lines: comparison of observed gravity change with observed vertical motion and with GIA models. *Report*, *Nordiska Kommissionen för Geodesi*, 2010. URL https://www.nordicgeodeticcommission.com/wp-cont

ent/uploads/2014/10/1-Makinen_et_al_land_uplift_gravity_lines.pdf. Accessed 22nd September, 2021. 335

S. Märdla. *Regional Geoid Modelling by the Least Squares Modified Hotine Formula Using Gridded Gravity Disturbances*. PhD thesis, Tallinn University of Technology, 2017. URL https://digikogu.taltech.ee/en/Item/baba5f3f-22ce-43b4-8f81-7b1fe015ac19. Accessed 17th September, 2021. 130, 218

P. J. Melchior. *The Tides of the Planet Earth*. Pergamon Press, Oxford, 1978. ISBN 978-0-0802-6248-2. 389, 392

J. X. Mitrovica, M. E. Tamisiea, J. L. Davis, and G. A. Milne. Recent mass balance of polar ice sheets inferred from patterns of global sea level change. *Nature*, 409:1026–1029, February 2001. URL https://doi.org/10.1038/35059054. 346

M. S. Molodensky, V. F. Eremeev, and M. I. Yurkina. *Methods for the Study of the External Gravitational Field and Figure of the Earth*. Israel Program of Scientific Translations, Jerusalem, 1962. (Transl. from Russian). 129, 173, 223, 404

D. Monniaux. Autograv CG5, 2011. URL https://commons.wikimedia.org/wiki/File:Autograv_CG5_P1150838.JPG. © 2011 David Monniaux (GFDL). Accessed 13th May, 2019. 301

H. Moritz. *Advanced Physical Geodesy*. H. Wichmann Verlag, Karlsruhe, 1980. ISBN 978-3-87-907106-7. 404, 444

W. H. Munk. The U.S. Commission on Ocean Policy, Testimony, 18th April 2002. URL https://govinfo.library.unt.edu/oceancommission/meetings/apr18_19_02/munk_statement.pdf. Accessed 14th May, 2019. 353

N

R. S. Nerem, D. P. Chambers, C. Choe, and G. T. Mitchum. Estimating mean sea level change from the TOPEX and Jason altimeter missions. *Marine Geodesy*, 33(1):435, 2010. URL https://doi.org/10.1080/01490419.2010.491031. 354

NOAA, Ocean currents. How does the ocean affect climate and weather on land? NOAA Ocean Exploration and Research. URL

https://oceanexplorer.noaa.gov/facts/climate.html. Accessed 1st March, 2020. 341

O

P.-A. Olsson, K. Breili, V. Ophaug, H. Steffen, M. Bilker-Koivula, E. Nielsen, T. Oja, and L. Timmen. Postglacial gravity change in Fennoscandia – three decades of repeated absolute gravity observations. *Geophysical Journal International*, 217:1141–1156, 2019. URL https://doi.org/10.1093/gji/ggz054. 335

P

R. L. Parker. The rapid calculation of potential anomalies. *Geophysical Journal of the Royal Astronomical Society*, 31:447–455, 1972. URL https://doi.org/10.1111/j.1365-246X.1973.tb06513.x. 252

M. K. Paul. A method of evaluating the truncation error coefficients for geoidal height. *Bulletin Géodésique*, 110:413–425, 1973. URL https://doi.org/10.1007/BF02521951. 224

N. K. Pavlis, S. A. Holmes, S. C. Kenyon, and J. K. Factor. The development and evaluation of the Earth Gravitational Model 2008 (EGM2008). *Journal of Geophysical Research*, 117(B4), 2012. URL https://doi.org/10.1029/2011JB008916. 77

W. R. Peltier. Home page, W. R. Peltier, FRSC. URL https://www.atmosp.physics.utoronto.ca/~peltier/data.php. Accessed 22nd September, 2021. 342

W. R. Peltier. Closure of the budget of global sea level rise over the GRACE era: the importance and magnitudes of the required corrections for global glacial isostatic adjustment. *Quaternary Science Reviews*, 28(17–18): 1658–1674, 2009. URL https://doi.org/10.1016/j.quascirev.2009.04.004. Special issue: Quaternary Ice Sheet-Ocean Interactions and Landscape Responses. 342

U. Pesonen. *Relative Bestimmungen der Schwerkraft in Finnland in den Jahren 1926–1929*. Publication 13, Finnish Geodetic Institute, Helsinki, 1930. 300

PIOMAS. Polar Science Center, PIOMAS Arctic Sea Ice Volume Reanalysis. URL http://psc.apl.washington.edu/research/projects/arctic-sea-ice-volume-anomaly/. Accessed 11th May, 2019. 375

M. Poutanen, M. Vermeer, and J. Mäkinen. The Permanent Tide in GPS Positioning. *Journal of Geodesy*, 70:499–504, 1996. URL https://doi.org/10.1007/BF00863622. 395

J. H. Pratt. II. On the attraction of the Himalaya mountains, and of the elevated regions beyond them, upon the plumb-line in India. *Philosophical Transactions of the Royal Society of London*, 145:53–100, 1855. URL https://doi.org/10.1098/rstl.1855.0002. 145

J. H. Pratt. II. On the deflection of the plumb-line in India caused by the attraction of the Himalaya mountains and the elevated regions beyond, and its modification by the compensating effect of a deficiency of matter below the mountain mass. *Proceedings of the Royal Society of London*, 9:493–496, 1859. URL https://doi.org/10.1098/rspl.1857.0096. 145

J. H. Pratt. On the degree of uncertainty which local attraction, if not allowed for, occasions in the map of a country, and in the mean figure of the earth as determined by geodesy ; a method of obtaining the mean figure free from ambiguity by a comparison of the Anglo-Gallic, Russian, and Indian Arcs ; and speculations on the constitution of the earth's crust. *Proceedings of the Royal Society of London*, 13:253–276, 1864. URL https://doi.org/10.1098/rspl.1863.0061. 145

K. Predehl, G. Grosche, S. M. Raupach, S. Droste, O. Terra, J. Alnis, Th. Legero, T. W. Hänsch, Th. Udem, R. Holzwarth, and H. Schnatz. A 920 km Optical Fiber Link for Frequency Metrology at the 19th Decimal Place. *Science*, 27th April 2012. URL https://doi.org/10.1126/science.1218442. 187

I. Prutkin. Gravitational and magnetic models of the core-mantle boundary and their correlation. *Journal of Geodynamics*, 45:146–153, 2008. URL https://www.researchgate.net/publication/257097255_Gravitational_and_magnetic_models_of_the_core-mantle_boundary_and_their_correlation. Accessed 1st March, 2020. 155

R

R. H. Rapp. The decay of the spectrum of the gravitational potential and the topography for the Earth. *Geophysical Journal International*, 99(3):449–455, 1989. URL https://doi.org/10.1111/j.1365-246X.1989.tb02031.x. 287

J. Richer. Observations astronomiques et physiques faites en l'isle de Caïenne. In *Ouvrages de Mathematique de M. Picard*. P. Gosse and J. Neaulme, 1731. URL http://www.e-rara.ch/zut/content/pageview/815403. Accessed 11th May, 2019. 300

R. A. Rohde. Post-glacial sea level rise, 2005. URL https://commons.wikimedia.org/wiki/File:Post-Glacial_Sea_Level.png. © 2005 Robert A. Rohde (GFDL). Accessed 21st July, 2019. 345

C. F. Ropelewski and P. D. Jones. An extension of the Tahiti-Darwin Southern Oscillation Index. *Monthly Weather Review*, 115:2161–2165, 1987. URL https://doi.org/10.1175/1520-0493(1987)115<2161:AEOTTS>2.0.CO;2. Accessed 13th January, 2022. 354

R. Rummel and F. Sansó, editors. *Satellite Altimetry in Geodesy and Oceanography. Proceedings, International Summer School of Theoretical Geodesy*, Lecture Notes in Earth Sciences, 50, Trieste, Italy, 25th May – 6th June 1992. Springer-Verlag. URL https://doi.org/10.1007/BFb0117924. 340

H. Ruotsalainen. Interferometric water level tilt meter development in Finland and comparison with combined earth tide and ocean loading models. *Pure and Applied Geophysics*, 2017. URL https://doi.org/10.1007/s00024-017-1562-6. 394

S

O. Sacks. Henry Cavendish: An early case of Asperger's syndrome? *Neurology*, 57(7):1347–1347, 2001. URL https://doi.org/10.1212/WNL.57.7.1347. 4

E. J. O. Schrama. *The Role of Orbit Errors in Processing of Satellite Altimeter Data*. PhD thesis, Delft University of Technology, 1989. URL https://www.ncgeo.nl/downloads/33Schrama.pdf. Accessed 11th May, 2019. 365

A. R. Schweiger, R. Lindsay, L. Zhang, M. Steele, H. Stern, and R. Kwok. Uncertainty in modeled Arctic sea ice volume. *Journal of Geophysical Research*, 116(C00D06), 2011. URL http://dx.doi.org/10.1029/2011JC007084. 375

D. Scuka. GOCE burning: Last orbital view. ESA blog, 2013. URL https://blogs.esa.int/rocketscience/2013/11/11/goce-burning-last-orbital-view/. Accessed 22nd September, 2021. 380

Sea Level Research Group. University of Colorado. URL https://sealevel.colorado.edu/. Accessed 22nd September, 2021. 354

J. Segawa, H. Fujimoto, and S. Okubo, editors. *Proceedings, IAG International Symposium on Gravity, Geoid and Marine Geodesy (GraGeoMar96)*, International Association of Geodesy Symposia 117, Tokyo, Japan, 30th September – 5th October 1996. Springer-Verlag. 461, 463

SourceForge, Maxima. Maxima, a computer algebra system. URL https://maxima.sourceforge.io/. Accessed 22nd September, 2021. 360

G. Spada and D. Melini. SELEN: a program for solving the "Sea Level Equation". User manual for version 2.9, Computational Infrastructure for Geodynamics (CIG), 2015. URL http://geodynamics.org/cig/software/selen/selen-manual.pdf. Accessed 1st March, 2020. 342

G. Strang van Hees. Stokes' Formula Using Fast Fourier Techniques. *Manuscripta geodaetica*, 15:235–239, 1990. 238

T

W. Torge. *Gravimetry*. de Gruyter, Berlin, New York, 1989. ISBN 978-3-11-010702-9. 404

W. Torge. Gravity and tectonics. In Kakkuri (1993), pages 131–172. 398

C. C. Tscherning and R. H. Rapp. Closed covariance expressions for gravity anomalies, geoid undulations, and deflections of the vertical implied by anomaly degree variances. *Report* 208, Dept. of Geodetic Science and Surveying, The Ohio State University, Columbus, OH, USA, 1974. URL https://earthsciences.osu.edu/sites/earthsciences.osu.edu/files/report-208.pdf. Accessed 17th January, 2020. 287

Tytyri Mine Experience. URL https://www.tytyrielamyskaivos.fi/en/mine-experience. Accessed 17th September, 2021. 394

V

P. Vaníček and E. Krakiwsky. *Geodesy – The Concepts*. Elsevier Science Publishers, second edition, 1987. ISBN 978-0-4448-7777-2. 404

F. A. Vening Meinesz. Gravity survey by submarine via Panama to Java. *The Geographical Journal*, 71(2):144–156, 1928. URL https://doi.org/10.2307/1782700. 300

M. Vermeer. Chronometric levelling. Report 83:2, Finnish Geodetic Institute, 1983a. 186, 313

M. Vermeer. A new Seasat altimetric geoid for the Baltic. Report 83:4, Finnish Geodetic Institute, 1983b. 351

M. Vermeer. *Geoid studies on Finland and the Baltic*. PhD thesis, University of Helsinki, 1984. Report 84:3, Finnish Geodetic Institute. 34, 113

M. Vermeer. FGI studies on satellite gravity gradiometry. 3. Regional high resolution geopotential recovery in geographical coordinates using a Taylor expansion FFT technique. Report 92:1, Finnish Geodetic Institute, 1992. 243

M. Vermeer. Geoid determination using frequency domain techniques. In Kakkuri (1993), pages 183–200. 234

M. Vermeer and J. Ádám, editors. *Proceedings, Second Continental Workshop on the Geoid in Europe*, Report 98:4, Finnish Geodetic Institute, Masala, 10th – 14th March, 1998. 462, 475

H. Virtanen. On superconducting gravimeter observations above 8 mHz at the Metsähovi station. Report 98:5, Finnish Geodetic Institute, Masala, 1998. 315

H. Virtanen. *Studies of Earth dynamics with superconducting gravimeter*. PhD thesis, University of Helsinki, 2006. URL http://urn.fi/URN:ISBN:952-10-3057-7. Publication 133, Finnish Geodetic Institute. Accessed 11th May, 2019. 316

H. Virtanen and J. Kääriäinen. The installation and first results from the superconducting gravimeter GWR20 at the Metsähovi station, Finland. Report 95:1, Finnish Geodetic Institute, Helsinki, 1995. 315

J. Vondrák, C. Ron, and V. Štefka. Earth orientation parameters based on EOC-4 astrometric catalog. *Acta Geodynamica et Geomaterialia*, 7(3):245–251, 2010. URL https://www.irsm.cas.cz/materialy/acta_content/2010_03/2_Vondrak.pdf. Accessed 20th February, 2020. 394

P. Vu, F. Frappart, J. Darrozes, V. Marieu, F. Blarel, G. Ramillien, P. Bonnefond, and F. Birol. Multi-satellite altimeter validation along the French Atlantic Coast in the Southern Bay of Biscay from ERS-2 to SARAL. *Remote Sensing*, 93(10), 2018. URL http://dx.doi.org/10.3390/rs10010093. 372

W

D. Watts. Fourifier, 2004. URL https://ejectamenta.com/imaging-experiments/fourifier/. Accessed 17th September, 2021. 246

L. Wen and D. L. Anderson. Layered mantle convection: A model for geoid and topography. *Earth and Planetary Science Letters*, 146(3–4):367–377, 1997. ISSN 0012-821X. URL http://222.195.83.195/wen/Reprints/WenAnderson97EPSL.pdf. Accessed 4th March, 2020. 155

H.-G. Wenzel. Ultra high degree geopotential model GPM3E97A to degree and order 1800 tailored to Europe. In Vermeer and Ádám (1998), pages 71–80. 82

P. Wessel, W. Smith, R. Scharroo, J. Luis, and F. Wobbe. Generic Mapping Tools: Improved Version Released. *EOS Trans. AGU*, 94(45):409–410, 2013. URL http://dx.doi.org/10.1002/2013EO450001. ii

K. Wieczerkowski, J. X. Mitrovica, and D. Wolf. A revised relaxation-time spectrum for Fennoscandia. *Geophysical Journal International*, 139:69–86, 1999. URL https://doi.org/10.1046/j.1365-246X.1999.00924.x. 346

Wikipedia, The Aerospace Corporation. URL https://en.wikipedia.org/wiki/The_Aerospace_Corporation. Accessed 23rd April, 2019. 221

Wikipedia, Declination. URL https://en.wikipedia.org/wiki/Declination. Accessed 29th February, 2020. 387

Wikipedia, Dislocation. URL https://en.wikipedia.org/wiki/Dislocation. Accessed 23rd April, 2019. 307

Wikipedia, Double-slit experiment. URL https://en.wikipedia.org/wiki/Double-slit_experiment. Accessed 18th February, 2020. 312

Wikipedia, Earth normal modes. URL https://en.wikipedia.org/wiki/Seismic_wave#Normal_modes. Accessed 16th March, 2020. 316

Wikipedia, Hour angle. URL https://en.wikipedia.org/wiki/Hour_angle. Accessed 29th February, 2020. 387

Wikipedia, John Pratt. URL https://en.wikipedia.org/wiki/John_Pratt_(Archdeacon_of_Calcutta). Accessed 23rd April, 2019. 145

Wikipedia, Mu-metal. URL https://en.wikipedia.org/wiki/Mu-metal. Accessed 23rd April, 2019. 315

Wikipedia, Pendulum clock. URL https://en.wikipedia.org/wiki/Pendulum_clock. Accessed 23rd April, 2019. 299

Wikipedia, Saros. URL https://en.wikipedia.org/wiki/Saros_(astronomy). Accessed 23rd April, 2019. 331

Wikipedia, Sea level rise. URL https://en.wikipedia.org/wiki/Sea_level_rise. Accessed 23rd April, 2019. 341

Wikipedia, Seasat conspiracy theory. URL https://en.wikipedia.org/wiki/Seasat#Conspiracy_theory. Accessed 23rd April, 2019. 351

Wikipedia, Strengthening mechanisms of materials. URL https://en.wikipedia.org/wiki/Strengthening_mechanisms_of_materials. Accessed 10th May, 2019. 307

Wikipedia, Sverdrup. URL https://en.wikipedia.org/wiki/Sverdrup. Accessed 23rd April, 2019. 339

Wikipedia, Zero-length springs. URL https://en.wikipedia.org/wiki/Spring_(device)#Zero-length_springs. Accessed 23rd April, 2019. 304

Wolfram Demonstrations, Difference formula for cosine. URL https://demonstrations.wolfram.com/DifferenceFormulaForCosine/. Accessed 7th April, 2019. 197

Wolfram Functions, $\sum_{k=1}^{\infty} \frac{\cos kx}{k}$. URL https://functions.wolfram.com/ElementaryFunctions/Cos/23/02/0001/. Accessed 25th February, 2020. 198

Wolfram MathWorld, Spherical Harmonic Addition Theorem. URL https://mathworld.wolfram.com/SphericalHarmonicAdditionTheorem.html. Accessed 11th May, 2019. 387

L. Wong and R. Gore. Accuracy of geoid heights from modified Stokes kernels. *Geophysical Journal of the Royal Astronomical Society*, 18(1):81–91, 1969. URL https://doi.org/10.1111/j.1365-246X.1969.tb00264.x. 221

G. Wöppelmann, C. Letetrel, A. Santamaría-Gómez, M.-N. Bouin, X. Collilieux, Z. Altamimi, S. D. Williams, and B. Martín-Míguez. Rates of sea-level change over the past century in a geocentric reference frame. *Geophysical Research Letters*, 36(12), 2009. URL https://doi.org/10.1029/2009GL038720. 333

Y

YouTube, Hammer vs. Feather, 2010. URL https://www.youtube.com/watch?feature=player_embedded&v=KDp1tiUsZw8#! Accessed 7th April, 2019. 4

D.-N. Yuan, W. L. Sjogren, A. S. Konopliv, and A. B. Kucinskas. Gravity field of Mars: A 75th degree and order model. *Journal of Geophysical Research*, 106 (E10):23 377–23 401, 2001. URL https://doi.org/10.1029/2000JE001302. 287

Index

Z